DESCRIPCIÓN Y MODELADO MEDIANTE REDES DE PETRI DE LOS PARÁMETROS Y COSTO-BENEFICIO DE UN SISTEMA DE COLAS EN ATENCIÓN AL PÚBLICO

CRISTIAN ANDRÉS LÓPEZ OSORIO – 1088274663

DARWIN SALAZAR GUERRERO – 1088305905

UNIVERSIDAD TECNOLÓGICA DE PEREIRA

FACULTAD DE INGENIERÍAS

PROGRAMA DE INGENIERÍA ELÉCTRICA

PEREIRA/RISARALDA

2019

DESCRIPCIÓN Y MODELADO MEDIANTE REDES DE PETRI DE LOS PARÁMETROS Y COSTO-BENEFICIO DE UN SISTEMA DE COLAS EN ATENCIÓN AL PÚBLICO

CRISTIAN ANDRÉS LÓPEZ OSORIO – 1088274663

DARWIN SALAZAR GUERRERO – 1088305905

Trabajo de grado presentado como requisito para optar por el título de Ingeniero Electricista

Director:

Mauricio Holguín Londoño

UNIVERSIDAD TECNOLÓGICA DE PEREIRA

FACULTAD DE INGENIERÍAS

PROGRAMA DE INGENIERÍA ELÉCTRICA

PEREIRA/RISARALDA

2019

DEDICATORIA

Darwin Salazar Guerrero

A mis padres Gloria Cecilia y Carlos Arturo quienes a pesar de las dificultades y en situaciones de escasez siempre con su entereza y tesón tan característico de su personalidad, me enseñaron que a pesar de todo, la educación siempre es importante.

Cristian Andrés López Osorio

No podría dedicar más este proyecto que a Dios, quien lo propició todo para que se hiciera realidad dándome el apoyo de mi familia en especial el de mi madre Aracelly Osorio Aguirre, quien me apoyó y me sigue apoyando en todo lo que emprendo en mi vida. Más que aprender sobre una ciencia, él me permitió ver que de su mano todo es posible y que las cosas más valiosas de la vida las enseña directamente él, por eso este triunfo va totalmente dedicado a Dios que cumple sus promesas en el tiempo perfecto.

AGRADECIMIENTOS

Darwin Salazar Guerrero

A mis padres Gloria Cecilia y Carlos Arturo que con mucho esfuerzo y dedicación, me apoyaron incondicionalmente para lograr una nueva meta en mi vida.

Al profesor Mauricio Holguín Londoño por su paciencia y conocimientos para salir con este proyecto adelante.

Cristian Andrés López Osorio

Agradezco a la Fundación Volar que a través del programa Becas Talento de la Universidad Tecnológica de Pereira me apoyaron en la mayor parte de mi carrera.

TABLA DE CONTENIDO

Pág.

TABLA DE FIGURAS

Pág.

LISTA DE TABLAS

Pág.

PARTE I

INTRODUCCIÓN

CAPÍTULO 1.

PLANTENAMIENTO DEL PROBLEMA

Las filas que se generan en lugares de atención al público son muy comunes en el diario vivir. Estas filas se presentan cuando un número de personas necesitan ser atendidas por un recurso compartido, es decir, varios clientes o usuario van a pagar a una misma caja, varios pacientes tienen consulta con el mismo médico o varias personas tienen que ser atendidas en oficinas de atención al público; estos recursos, de forma genérica, conocen como servidores [1].

Estas filas no están lejos de ninguna persona, por ejemplo, cuando se disponen a abordar el transporte público cada persona debe hacer fila hasta que llegue el turno de subir y pagar el pasaje, posteriormente puede ser que tengan que esperar por un puesto libre y esto sometido al tiempo de llegada a su destino. Y así como en el ejemplo anterior, se encuentran innumerables sistemas donde hay líneas de espera y su comportamiento suele ser más complejo como en los mismos sistemas de atención al público, donde se pueden encontrar novedades como colas que se pueden diferenciar, ejemplo, en un banco hay una fila preferencial donde se atiende X tipo de clientes y fila general donde se atienden Y tipos de clientes y cada fila puede tener una distribución de llegada y tiempos de servicios diferentes e incluso no deterministas [1], [2].

Los parámetros estructurales en líneas de espera son: el número de servidores en el sistema, longitud de la cola, tiempo en el sistema y tiempo de espera en la cola. Estos parámetros se ven afectados negativamente cuando la demanda es mayor a la capacidad del servicio, provocando así largas colas, tiempo de espera excesivo, saturación y bloqueos en el sistema o también cuando la capacidad es mayor a la demanda donde los servidores pueden ser subutilizados [3].

El análisis de las líneas de espera es complejo ya que el control de los parámetros se ve afectado por distribuciones que pueden ser estocásticas y su solución intuitiva tiene un margen de error alto, pero no solucionarlos es un riesgo para cualquier empresa [3].

Por lo descrito, se sabe que el tiempo, la cantidad de servidores, la capacidad de la cola y la capacidad de un recurso acarrean consecuencias tales como toma de decisiones, afectación del correcto funcionamiento del sistema, malestar de los usuarios, entre otros. Por lo tanto, se busca un método de análisis de líneas de espera que permita su estudio y tomas de decisiones relacionadas con los tiempos

de la cola, que garantice la continuidad del sistema sin que se sature y permita el análisis entre el costo de un servicio y su beneficio.

CAPÍTULO 2.

JUSTIFICACIÓN

En los sistemas de atención al cliente se pueden definir por índices de calidad, como el tiempo que demora un usuario en recibir un servicio y el cumplimiento de las leyes que le regula; así lo perciben las organizaciones de atención al público hoy en día y se han convertido en una necesidad independiente de donde se encuentren, quien sean sus clientes o cuál sea su dimensión. Es por ello por lo que el estudio de estos sistemas es de suprema importancia para las empresas [4].

La teoría de colas, de autoría del matemático danés Agner Krarup Erlang de 1909 [5], ha evolucionado en estudio de los sistemas de líneas de espera con base en modelos de distribución probabilística. Esta teoría ofrece un método para describir fácil y definitivamente las filas en términos matemáticos, así se puede analizar las ecuaciones para comprender y predecir el comportamiento de la cola. Gracias a los avances computacionales, la teoría de colas ha resuelto algunas de sus desventajas a través de la simulación de eventos discretos, ya que puede verse como una secuencia de estados delimitados por eventos que ocurren de manera estocástica [5].

El concepto de eventos discretos tiene por finalidad identificar a sistemas en los que los eventos, que cambian el estado del mismo, ocurren en instantes espaciados en el tiempo [6]. Ejemplo de ello es el tiempo de llegada entre personas a una fila de un banco o de vehículos esperando en una caseta de cobro. Como se puede observar en los ejemplos anteriores, el valor de sus variables cambia en momentos de tiempo no necesariamente constantes; en particular, la cantidad de personas que están en un banco puede ir aumentando conforme los clientes llegan, pero va a haber lapsos de tiempo que se mantiene u otros instantes de tiempo donde la cantidad de personas disminuye.

Ahora bien, las Redes de Petri (RdP) son un formalismo también muy utilizado en el modelamiento y el análisis de sistemas de eventos discretos (SED). Esta popularidad se debe a que combinan un sólido fundamento matemático, representación gráfica y capacidad de modelar procesos paralelos y distribuidos. Se considera a las RdP como herramienta que proveen un ambiente uniforme para el modelado, análisis formal y diseño de SED. Las Redes de Petri pueden ser usadas para el análisis del comportamiento, evaluación de acciones, simulación y construcción de controladores [7].

Matemáticamente, una RdP puede ser descrita por un conjunto de ecuaciones que reflejen el comportamiento del sistema. Esto permite realizar un análisis formal, el cual, consiste en examinar dichas ecuaciones y sus propiedades, relacionándolas con eventos tangibles, como son: operaciones concurrentes, liberaciones de bloqueo, apropiada sincronización, actividades repetitivas, exclusiones mutuas, operaciones con recursos compartidos, etc. [7]

Las RdP han sido utilizadas para el análisis y solución de sistemas discretos en donde se presentan líneas de espera, como es el caso de los sistemas de manufactura descritos en [8], donde se propone una metodología basada en RdP para realizar el análisis, modelamiento y la simulación de los procesos, permitiendo modificar los parámetros de la red para poder determinar si los cambios son económicamente viables.

Dado lo anterior, el presente documento plantea realizar el análisis de los parámetros y el estudio del costo-beneficio para un sistema de colas aplicado a la atención de público mediante modelamiento y descripción por Redes de Petri.

CAPÍTULO 3.

OBJETIVOS

3.1. OBJETIVO GENERAL

Describir y modelar, mediante Redes de Petri, los parámetros que definen la configuración de un sistema de colas y su costo-beneficio, aplicado a la atención de público en cajas de pago utilizando modelos de distribución probabilística.

3.2. OBJETIVOS ESPECÍFICOS

- Indagar los conceptos fundamentales acerca de la descripción de la teoría de colas y sus parámetros descriptivos.
- Profundizar en los conceptos fundamentales acerca de la teoría de Redes de Petri y su uso para la descripción de sistemas discretos.
- Establecer un método para describir y modelar sistemas de colas mediante Redes de Petri.
- Aplicar el modelo al análisis de un sistema de atención al público en cajas de pago.

PARTE II.

MARCO CONCEPTUAL

CAPÍTULO 4.

DISTRIBUCIONES PROBABILÍSTICAS

4.1. DISTRIBUCIÓN EXPONENCIAL

El análisis de los distintos modelos de colas está determinado en gran parte por la distribución de probabilidad de los tiempos entre llegadas y la distribución de los tiempos de servicio. En los sistemas de colas reales, estas distribuciones pueden tomar prácticamente cualquier forma. Sin embargo, para formular y analizar un modelo matemático es necesario especificar la forma supuesta de cada una de estas distribuciones. Para que sea útil, la forma expuesta debe ser lo suficientemente realista como para que el modelo entregue predicciones razonables y al mismo tiempo lo suficientemente manejable para que sean posibles tales predicciones. Con estas consideraciones, la distribución exponencial es la distribución de probabilidad más importante en la teoría de colas [9].

- **Variable aleatoria exponencial**

Es una de las variables aleatorias continuas más sencillas de definir, dado que se caracteriza únicamente con un parámetro y su formulación también presenta ventajas desde el punto de vista analítico, al resultar cómodo realizar operaciones con ella [10].

Una variable aleatoria continua X tiene una distribución exponencial si su función de densidad f_X tiene la siguiente expresión:

$$f_X(x) = \lambda e^{-\lambda x}, \quad x \geq 0$$

Donde $\lambda > 0$ es el único parámetro que caracteriza la variable aleatoria.

En la Figura 1 se representa la función densidad de una variable aleatoria exponencial para distintos valores del parámetro. Se observa que la función de densidad toma el valor de λ en el eje de ordenadas $(x = 0)$, por lo que un valor elevado en dicho punto supone que la variable tenderá a generar valores relativamente bajos y viceversa.

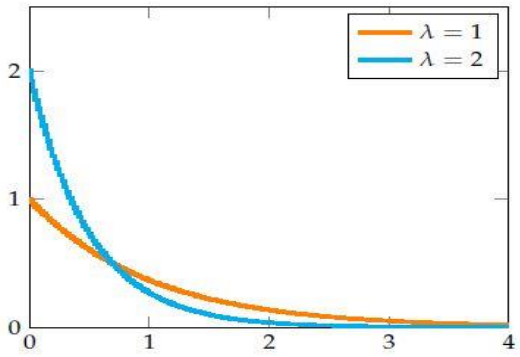


Figura 1. Función de densidad de probabilidad de una variable aleatoria exponencial *[10]*.

Para calcular la esperanza se aplica la definición de la misma para el caso continuo:

$$\mathbb{E}[X] = \int_x xf(x)\,dx$$

Lo que se traduce para el caso de la variable aleatoria exponencial en

$$\mathbb{E}[X] = \int_0^{-\infty} x\lambda e^{-\lambda x}\,dx$$

Dicha integral se resuelve por partes, lo que lleva a

$$\mathbb{E}[X] = \left| -xe^{-\lambda x} - \frac{1}{\lambda}e^{-\lambda x} \right|_0^{-\infty} = \frac{1}{\lambda}$$

- **La propiedad "sin memoria"**

Se dice que una variable aleatoria X no tiene memoria si cumple la siguiente propiedad

$$P_r(X > t + s | X > t) = e^{-\lambda s} = P_r(X > s) \qquad (1)$$

Para el caso en que X represente un tiempo de vida, la propiedad sin memoria supone que la probabilidad de que sobreviva un tiempo $t + s$, dado que ya ha sobrevivido un tiempo t, es igual a la probabilidad de que sobreviva un tiempo s partiendo desde 0 (es decir, como si acabase de ser puesto en funcionamiento). Esto es, que la distribución del tiempo restante de vida no depende del tiempo t que lleve con vida [10].

La variable aleatoria exponencial no tiene memoria, como se puede demostrar desarrollando la expresión de la parte izquierda de (1):

$$P_r(X > s + t | X > t) = \frac{p_r(X > s + t, X > t)}{P_r(X > t)}$$

En el cociente, la probabilidad de $X > t$ del numerador ya está incluida en $X > t + s$, por lo que la parte izquierda de (1) se puede obtener como:

$$P_r(X > s + t | X > t) = \frac{P_r(X > s + t)}{P_r(X > t)} = \frac{e^{-\lambda(s+t)}}{e^{-\lambda t}} = e^{-\lambda s} \quad (2)$$

Que corresponde con la parte derecha de (1):

$$P_r(X > s) = e^{-\lambda s}$$

- **Mínimo de variables aleatorias exponenciales**

El mínimo de variables aleatorias exponenciales independientes sigue una distribución exponencial. Sean $X_1, \cdots, X_n$ variables aleatorias independientes con distribuciones exponenciales de parámetros $\lambda_1, \cdots, \lambda_n$, respectivamente, entonces, la variable aleatoria $X = mín\{X_1, \cdots, X_n\}$ sigue una distribución exponencial de parámetro $\lambda = \lambda_1 + \cdots + \lambda_n$.

4.2. PROCESOS DE POISSON

Un proceso de conteo representa el número total de ocurrencias de un determinado evento hasta dicho instante de tiempo. Hay infinidad de ejemplos de procesos de conteo: número de veces que se reinicia un computador, *Champions League* que ha ganado un equipo de fútbol, visitas totales que tiene un video en *YouTube*, etc. [10].

$\{X_n, n \geq 1\}$ es una colección de variables aleatorias que representan tiempos entre sucesos. Se define como

$$S_0 = 0 \qquad S_n = X_1 + \cdots + X_n$$

S_n es el tiempo en el que ocurre el *n*-ésimo suceso. Sea

$$N(t) = máx\{n \geq 0 | S_n \leq t\}, \text{ para todo } t \geq 0$$

$N(t)$ es el número de sucesos ocurridos en el intervalo $(0, t]$. $\{N(t), t \geq 0\}$ se denomina ***proceso de conteo***.

Cabe notar que $N(0) = 0$, esto quiere decir que no hay eventos antes del instante inicial y $N(t)$ salta con pasos de longitud 1 en los tiempos $t = S_n$, $n = 1,2, \ldots$

Algunas propiedades que satisface un proceso de conteo son las siguientes:

- $N(t)\, \epsilon\, \mathbb{Z}_+, \ \forall t \geq 0$.
- Si $s < t$, entonces $N(s) \leq N(t)$ y $N(s) - N(t)$ es el número de sucesos que han ocurrido en el intervalo $(s, t]$.

Si $\{X_n, n \geq 1\}$ es una sucesión de variables aleatorias independientes con idéntica $\text{Exp}\ (\lambda)$, entonces $\{N(t), t \geq 0\}$ se denomina ***proceso de Poisson*** de parámetro λ y se representa por $PP(\lambda)$.

Se observa que un proceso de Poisson (como en general, un proceso de conteo) es un proceso estocástico continuo con espacio de estados discreto. A continuación, se pueden ver un par de propiedades sobre el comportamiento de $\{N(t), t \geq 0\}$.

Sea $t \geq 0$ fijo, entonces

$$P(N(t) = k) = e^{-\lambda t} \frac{(\lambda t)^k}{k!} \tag{3}$$

Para hacer la demostración de la ecuación (3), se describe que:

- $N(t) \geq k \Leftrightarrow k$ o más sucesos ocurren durante $(0, t]$
- el k-ésimo suceso tiene lugar en t o antes
- $S_k \leq t$

Luego,

$$P(N(t) \geq k) = P(S_k \leq t),$$

De donde se sigue que

$$\begin{aligned} P(N(t) = k) &= P(N(t) \geq k) - P(N(t) \geq k + 1) \\ &= P(S_k \leq t) - P(S_{k+1} \leq t) \end{aligned}$$

Por otro lado, puesto que S_k es una suma de variables aleatorias exponenciales con idéntica distribución, se tiene que

$$S_k \sim Gamma(k, \lambda)$$

Finalmente.

$$P(N(t) = k) = P(S_k \leq t) - P(S_{k+1} \leq t)$$

$$= \left[1 - \sum_{r=0}^{k-1} e^{-\lambda t} \frac{(\lambda t)^r}{r!}\right] - \left[1 - \sum_{r=0}^{k} e^{-\lambda t} \frac{(\lambda t)^r}{r!}\right]$$
$$= e^{-\lambda t} \frac{(\lambda t)^k}{k!}$$

Un proceso de Poisson cuenta con dos propiedades que a continuación se presentan:

- La propiedad de incrementos independientes significa que el número de sucesos producidos en intervalos de tiempo disyuntos son independientes. Por ejemplo, el número de sucesos que han ocurrido hasta el tiempo 10, $N(10)$, debe ser independiente del número de sucesos que ocurran en el intervalo $(10,15]$, $N(15) - N(10)$ [9].
- La propiedad de incrementos estacionarios significa que la distribución del número de sucesos que ocurren en cualquier intervalo de tiempo sólo depende de la longitud de dicho intervalo. En otras palabras, el número de sucesos ocurridos en el intervalo $(t_1 + s, t_2 + s]$, $N(t_2 + s) - N(t_1 + s)$, siguen la misma distribución que el número de sucesos ocurridos en el intervalo $(t_1, t_2]$ [9].

Para caracterizar un proceso de Poisson primero se necesita previamente la siguiente definición:

Una función $f: \mathbb{R} \to \mathbb{R}$ se dice que es $o(h)$ si

$$\lim_{h \to 0} \frac{f(h)}{h} = 0$$

Por ejemplo

- $f(x) = 2x$ no es $o(h)$, $f(x) = x^2 + 3x^3$ es $o(h)$
- $f(x) = e^{\lambda x} - 1 - \lambda x$ es $o(h)$
- Si $f(x)$ y $g(x)$ son $o(h)$, entonces $f(x) + g(x)$ es $o(h)$
- Si $f(x)$ es $o(h)$ y $c \in \mathbb{R}$, entonces $c * f(x)$ es $o(h)$.

Entonces un proceso de conteo $\{N(t), t \geq 0\}$ es un $PP(\lambda)$ si $N(0) = 0$ y además:

- $P(N(h) = 0) = 1 - \lambda h + o(h)$
- $P(N(h) = 1) = \lambda h + o(h)$
- $P(N(h) = j) = o(h)$, para todo $j \geq 2$

CAPÍTULO 5.

TEORÍA DE COLAS

El origen de la Teoría de Colas se debe al estudio de Agner Krarup Erlang (Longborg, Dinamarca, 1878 – 1929) en 1909 para la solución de la congestión de líneas telefónicas. El trabajo inicial de Erlang lo continuaron diversos investigadores en la primera mitad del siglo XX, como Pollaczek, Kolmogorov y Khintchine, entre otros. A partir de la década de los años 50 hubo un considerable crecimiento de esta área, y su interés ha ido yendo en aumento debido en parte, al gran desarrollo del ámbito de las telecomunicaciones, uno de los campos donde la Teoría de Colas tiene una mayor implicación [11].

5.1. ESTRUCTURA DE UN SISTEMA DE COLAS

El proceso básico de un modelo de colas normalmente se describe de la siguiente manera: Una *fuente de entrada* genera en el tiempo un *cliente* que requiere un servicio, estos clientes entran en un *sistema de colas* y se unen. Por medio de una regla específica, en determinado tiempo, se selecciona un miembro de la cola para ser servido, esta regla se llama *disciplina de la cola*. El servicio requerido es realizado por el cliente por un *mecanismo de servicio*, después de esto, el cliente abandona el sistema de colas [12]. La Figura 2 ilustra los componentes básicos de un sistema de colas.

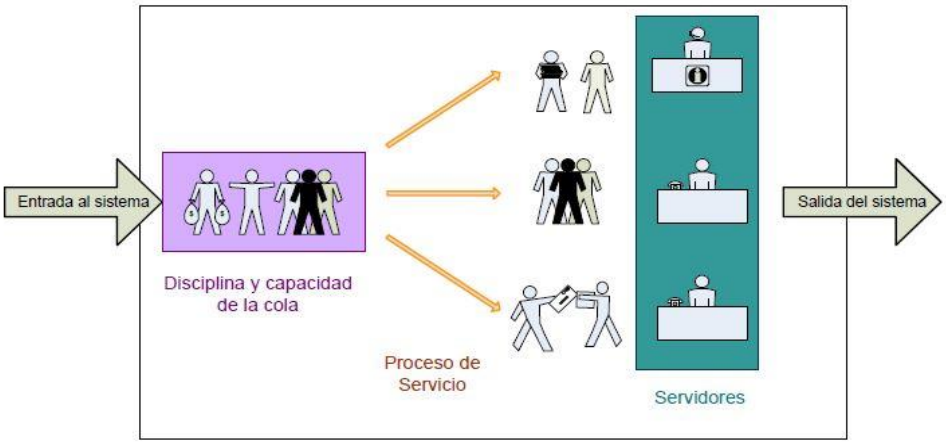


Figura 2. Componentes de un sistema de colas *[2]*.

5.2. CARACTERÍSTICAS DE LOS SISTEMAS DE COLAS

Las características básicas que se utilizan para describir un sistema de colas son las siguientes [13]:

- **Patrón de llegada de los clientes**

Es la forma en cómo llegan los clientes a solicitar un servicio. Al momento de referirse a la población de clientes es necesario especificar la distribución de probabilidad con la cual se generan estos clientes en el tiempo, en el caso que estas llegadas sean probabilísticas y no determinísticas (cuando los tiempos de llegada de clientes al servidor son constantes). Dada la naturaleza de generación de clientes (llegadas) en el sistema, lo normal es describir dicha generación como un proceso de Poisson, es decir, el número de clientes que llegan a un sistema en un momento específico se describe mediante una distribución de probabilidad [14].

- **Patrones de servicio de los servidores**

Los servidores pueden tener un tiempo de servicio variable, en cuyo caso hay que asociarle una función de probabilidad. También pueden atender en lotes o de modo individual [13].

El tiempo de servicio también puede variar con el número de clientes en la cola, trabajando más rápido o más lento [13]. Al igual que el patrón de llegadas, el patrón de servicios también se puede describir mediante una función de probabilidad.

- **Disciplina de la cola**

La disciplina de la cola es la manera en que los clientes se ordenan al momento de ser servidos. Cuando se piensa en colas se intuye que la disciplina de cola es normalmente FIFO (atender primero a quien llegó primero). Sin embargo, en muchas colas es habitual el uso de la disciplina LIFO (atender primero al último). También es posible encontrar reglas de secuencia con prioridades, como por ejemplo, primero las tareas con menor duración o según tipos de clientes [13].

- **Capacidad del sistema**

En algunos sistemas existe una limitación respecto al número de clientes que pueden esperar en la cola. A estos casos se les denomina situaciones de cola finitas. Esta limitación es una simplificación en la modelización de la impaciencia de los clientes o de la capacidad de retener clientes en fila [13].

- **Números de canales del servicio**

Evidentemente es preferible utilizar sistemas multiservidores con una única línea de espera para todos, que con una cola por servidor porque en un sistema con única cola es más fácil registrar el tiempo de llegada de los clientes. Por lo tanto, cuando se habla de canales de servicio paralelos, se habla generalmente de una cola que alimenta a varios servidores mientras que el caso de colas independientes se asemeja a múltiples sistemas con sólo un servidor [13]. En la Figura 3 se observa que el sistema que lleva como nombre *Variante 1* hace referencia a un sistema de

fila única con múltiples servidores mientras que el sistema *Variante 2* representa un sistema de varias colas con múltiples servidores.

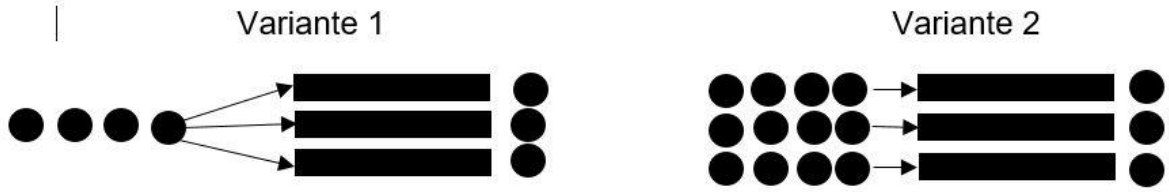


Figura 3. Sistemas de colas multicanal *[13]*.

- **Etapas de servicio**

Un sistema de colas puede ser unietapa o multietapa. En los sistemas multietapa el cliente puede pasar por un número de etapas mayor a uno. Por ejemplo, una peluquería es un sistema unietapa, pero si cuenta con diferentes servicios (manicura, maquillaje) cada uno de estos servicios se desarrolla por un servidor diferente, siendo entonces un sistema multietapa [13]. En algunos sistemas multietapa se puede admitir la vuelta atrás o "reciclado" como se muestra en la Figura 4, esto es habitual en sistemas productivos como controles de calidad y reprocesos [13].

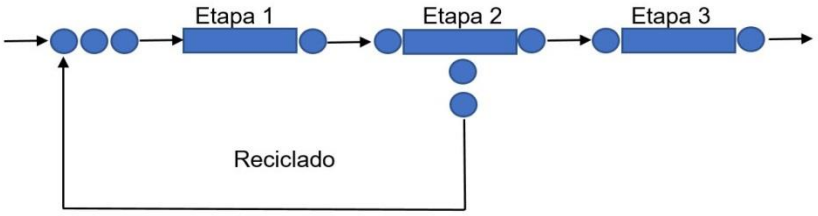


Figura 4. Sistema multietapa con retroalimentación *[13]*.

5.3. NOTACIÓN KENDALL

Con objeto de estandarizar la forma de definir los posibles sistemas de espera, se usa la *Notación Kendall*, la cual fue originalmente propuesta por el inglés David G. Kendall en 1953 con tres primeros parámetros los cuales se emplean para especificar las características de una cola [10]:

$$A/B/c$$

Estos parámetros representan:

- A: Especifica cómo se distribuye el tiempo entre peticiones, es decir, la variable aleatoria entre una petición que llega al sistema y la siguiente.

- B: Especifica cómo es la variable aleatoria del tiempo de servicio, esto es, el tiempo que pasa desde que una petición accede a un recurso hasta que es atendida. Tanto A como B suelen servir para indicar alguna de las siguientes distribuciones de tiempo:

  - M (por Markov, o "sin Memoria"): en este caso se trata de la variable aleatoria exponencial.

  - D: si se trata de un caso en que dicho tiempo es una constante (esto es, se trata de una variable determinista).

  - G: si se analiza un caso genérico, sin tener que especificar cómo se distribuye la variable aleatoria que determina alguno de los tiempos.

- c: número de recursos idénticos en paralelo.

Además de estos parámetros de la notación de Kendall, en ocasiones se emplean dos variables más, por lo que en el caso más completo la representación sería [10]:

$$A/B/c/K/N/Z$$

El significado de los dos últimos parámetros es el siguiente:

- K: capacidad máxima de todo el sistema, es decir, número máximo de peticiones que caben a la vez (en la cola y en los servidores).

- N: hace referencia al tamaño (finito o no) de la población, lo que determina cómo varía la tasa de peticiones en función de las que se encuentran en el sistema (por ejemplo, si un usuario no realiza una petición hasta que la anterior ha sido atendida).

- Z: especifica la disciplina de la cola, por ejemplo, si en vez de FIFO se emplea algún tipo de mecanismo que da prioridad a unos usuarios frente a otros.

Los tres primeros parámetros de la notación de Kendall aparecerán siempre, mientras que los tres últimos sólo suelen hacerse explícitos cuando no es suficiente con los valores por defecto: si la capacidad del sistema K y la población N son infinitos y si la disciplina de la cola es FIFO, no se suelen indicar.

5.4. COSIDERACIONES PARA EL ANÁLISIS DE LOS MODELOS DE COLAS

➢ Notación y terminología

La notación y terminología estándar que se utilizarán durante los siguientes apartados es la siguiente [15]:

- Estado del sistema = número de clientes en el sistema
- Longitud de la cola = número de clientes en cola = Estado del sistema – número de clientes en servicio.
- $n(t)$: número de clientes en el sistema en el instante t
- $p_n(t)$: probabilidad de que exactamente n clientes estén en el sistema en el instante t.
- s: número de servidores en el sistema
- λ_n: tasa media de llegadas (número esperado de llegadas por unidad de tiempo) de nuevos clientes cuando hay n clientes en el sistema.
- μ_n: tasa media de servicio para todo el sistema (número esperado de clientes que completan su servicio por unidad de tiempo) cuando hay n clientes en el sistema.
- $W_q^{(n)}$: es el tiempo de espera en cola del cliente n-ésimo.

Cuando λ_n es constante para todo n, se denota por λ. Esto significa que el número medio de clientes que llega al sistema por unidad de tiempo no depende del estado del sistema. Cuando la tasa media de servicio por servidor ocupado es constante, se denota por μ. En este caso, $\mu_n = s\mu$, cuando $n \geq s$, es decir, cuando los s servidores están ocupados. En estas circunstancias, $\frac{1}{\lambda}$ es el tiempo esperado entre llegadas, $\frac{1}{\mu}$ es el tiempo de servicio esperado y $\rho = \frac{\lambda}{s\mu}$ es el factor de utilización del sistema, es decir, la fracción media de tiempo que los servidores están ocupados [15].

➢ Estado transitorio vs estado estacionario

Cuando un sistema de colas inicia su operación, los distintos factores del sistema se encuentran bastante influenciados por las condiciones iniciales; se dice que el sistema se encuentra en estado transitorio. Una vez que ha pasado suficiente tiempo, usualmente, los factores del sistema se vuelven independientes de las

condiciones iniciales y del tiempo transcurrido, y se dice que el sistema se encuentra en estado estable [9].

Para obtener las probabilidades de estado, $p_n(t)$: probabilidad de que el sistema esté en el estado n (haya n clientes en el sistema) en el instante t, en general se debe calcular las probabilidades $p_n = \lim_{t\to\infty} p_n(t)$ cuando el sistema se encuentra en estado estable. Esto significa que se debe esperar a que el sistema se vuelva independiente de las condiciones iniciales y del tiempo que ha transcurrido desde el inicio del mismo. Por lo tanto, la probabilidad estacionaria p_n se puede interpretar como la probabilidad de que haya n clientes en el sistema cuando éste ha alcanzado el estado estacionario. Hay que puntualizar que no todos los sistemas tienen estado estacionario, pues $\lim_{t\to\infty} p_n(t)$ podría no dar lugar a una distribución de probabilidad [9].

5.5. MODELO DE COLAS DETERMINÍSTICO $D/D/1$

Este es considerado el caso más elemental de un modelo de colas, ya que tanto las tasas de llegada como las tasas de servicio son conocidas con exactitud. Los clientes son atendidos según una disciplina FIFO. Conocida la notación mostrada anteriormente, se analizará el modelo $D/D/1/K-1$, puesto que, si no se tiene control sobre el límite de la cola, se tendría un modelo en donde la cola crece de forma indefinida [16].

Donde

- $D/\cdot/\cdot/\cdot$ significa que el tiempo entre llegadas es constante e igual a $\frac{1}{\lambda}$, $\lambda > 0$, entonces λ denota el número de llegadas por unidad de tiempo, llamado *tasa de llegadas*.

- $\cdot/D/\cdot/\cdot$ significa que el tiempo de servicio es constante e igual a $\frac{1}{\mu}$, $\mu > 0$, donde μ es el número de servicios por unidad de tiempo en período de ocupación, llamado *tasa de servicio*.

- $\cdot/\cdot/1/\cdot$ indica que hay solo un servidor.

- $\cdot/\cdot/\cdot/k-1$ se refiere a la *capacidad del sistema*, es decir, el número máximo de cliente que admite el sistema. Cuando un cliente está en el servicio, sólo puede haber k-2 clientes en cola, y el k-ésimo cliente que aspire a entrar al sistema será rechazado.

Para un correcto análisis se asume que $\lambda > \mu$, de lo contrario no se formaría ninguna cola, ya que cada cliente será atendido de manera inmediata.

Interesa conocer el estado del sistema en el instante t, es decir, conocer $n(t)$, $L(t)$ y el tiempo que afecta al cliente *n*, es decir, $W_q^{(n)}$.

Primero que todo, se asume que en el tiempo $t = 0$ no hay clientes $(N(0) = 0)$, se define a τ como el instante de tiempo en el que se produce el primer rechazo, es decir, llega un cliente cuando en el sistema ya hay k-1 clientes.

Si $t \in \left(0, \frac{1}{\lambda}\right)$, se dice que al sistema todavía no ha llegado ningún cliente.

Si $t \in \left(\frac{1}{\lambda}, \tau\right)$, entonces $n(t)$ es igual al número de llegadas menos el número de salidas hasta t, donde

$$\#\text{ llegadas hasta } t = \left[\frac{t}{\frac{1}{\lambda}}\right] = [\lambda t] \qquad \#\text{ salidas hasta } t = \left[\frac{t-\frac{1}{\lambda}}{\frac{1}{\mu}}\right] = \left[\mu t - \frac{\mu}{\lambda}\right]$$

Entonces

$$n(t) = [\lambda t] - \left[\mu t - \frac{\mu}{\lambda}\right] \tag{4}$$

Para $t \geq \tau$, el tiempo de servicio es un múltiplo entero del tiempo entre llegadas, es decir,

$$m\frac{1}{\lambda} = \frac{1}{\mu}$$

En este caso, siempre que ocurra la salida de un cliente hay una llegada simultánea. De esta manera, no puede ocurrir simultáneamente una salida y un rechazo, puesto que, si un cliente sale, deja un lugar libre en el sistema para el otro cliente que llega en ese momento. En consecuencia, el número de clientes es creciente hasta que a partir del instante τ es constante e igual a la capacidad del sistema, $K - 1$. Por lo tanto, se tiene

$$n(t) = \begin{cases} 0 & si \quad t < \frac{1}{\lambda}, \\ [\lambda t] - \left[\mu t - \frac{\mu}{\lambda}\right] & si \quad \frac{1}{\lambda} < t < \tau, \\ k - 1 & si \quad t \geq \tau \end{cases}$$

Como el sistema tiene disciplina FIFO, se definen los valores asociados a cada cliente, así:

$S^{(n)}$: es el tiempo del n-ésimo servicio.

$T^{(n)}$: es el tiempo entre la n-ésima y la (n+1)-ésima llegada aceptada.

Ahora, teniendo claro estos términos se puede expresar a $W_q^{(n+1)}$ en función de $W_q^{(n)}, S^{(n)}$ y $T^{(n)}$.

Si $W_q^{(n)} + S^{(n)} \leq T^{(n)}$ entonces $w_q^{(n+1)} = 0$.

Si $W_q^{(n)} + S^{(n)} \geq T^{(n)}$ entonces $w_q^{(n+1)} = W_q^{(n)} + S^{(n)} - T^{(n)}$.

Aunque esto es cierto en general, en el caso del sistema que se está analizando, si $t < \tau$, entonces:

$$S^{(n)} = \frac{1}{\mu} \quad T^{(n)} = \frac{1}{\lambda}$$

Por lo que

$$w_q^{(n+1)} = W_q^{(n)} + \frac{1}{\mu} - \frac{1}{\lambda}$$

Si $t \geq \tau$, entonces cualquier cliente que sea aceptado será servido después de que hayan sido atendidos los k-2 que le preceden (y no los k-1, pues su llegada ha coincidido con una salida). Así

$$W_q^{(n)} = \frac{k-2}{\mu}$$

En resumen,

$$w_q^{(n)} = \begin{cases} (n-1)\left(\frac{1}{\mu} - \frac{1}{\lambda}\right) & si \;\; n = 1,2,\dots,\lambda\tau - 1 \\ \frac{k-2}{\mu} & si \;\; n \geq \lambda\tau \end{cases} \quad (5)$$

5.6. MODELOS DE COLAS EXPONENCIALES CON UN ÚNICO SERVIDOR

5.6.1. Modelo $M/M/1$

Se trata de un sistema en el que el tiempo entre llegadas es exponencial, hay un único recurso para atender las peticiones, y la longitud de lo cola no está limitada. El $M/M/1$ es el sistema más básico y puede servir para modelar sistemas sencillos

con único recurso atendiendo a una población variada, como, por ejemplo, un punto de acceso *Wifi* que transmita los datos de varios usuarios [10].

El procedimiento de análisis a seguir se basa en tres pasos [9]:

1. Obtener las ecuaciones en diferencia para $p_n(t)$
2. Obtener las ecuaciones diferenciales en diferencia para $p_n(t)$
3. Obtener las probabilidades límite p_n para el comportamiento estacionario.

Para obtener las ecuaciones en diferencia para $p_n(t)$ se analiza cómo el sistema puede alcanzar el estado n en el instante $t + \Delta t$. Las posibilidades para ello son las siguientes [9]:

- Si en el instante t el sistema estaba en el estado n, entonces en $(t, t + \Delta t]$ se produjeron $j \geq 0$ llegadas y j servicios.
- Si en el instante t el sistema estaba en el estado $n + j$, entonces en $(t, t + \Delta t]$ se registraron k llegadas y $j + k$ servicios.
- Si en el instante t el sistema estaba en el estado $n - j$, entonces en $(t, t + \Delta t]$ se produjeron $j + k$ llegadas y k servicios.

Puesto que los tiempos de llegadas y servicio son exponenciales, se sabe que los procesos que rigen el número de servicios que se producen hasta un cierto instante son Procesos de Poisson, y por lo tanto la probabilidad de que se den dos o más llegadas o servicios en un intervalo de longitud Δt es $o(\Delta t)$. Por lo tanto, basta considerar los casos en los que a lo sumo se den una llegada y un servicio [9]. Por lo tanto, para cada $n \geq 1$, las posibilidades son las siguientes:

- En el instante t el sistema estaba en el estado n y
  - En $(t, t + \Delta t]$ no se produjeron ni llegadas ni servicios
  - En $(t, t + \Delta t]$ se produjeron 1 llegada y 1 servicio
- En el instante t el sistema estaba en el estado $n + 1$ y en $(t, t + \Delta t]$ no se registraron llegadas y sólo finalizó un servicio
- En el instante t el sistema estaba en el estado $n - 1$ y en $(t, t + \Delta t]$ se registró 1 llegada y no finalizó ningún servicio.

Por lo tanto,

$$p_n(t + \Delta t) = P(n\ clientes\ en\ t, no\ llegadas\ ni\ servicios\ en\ (t, t + \Delta t])$$

$$+ P(n\ clientes\ en\ t, una\ llegada + un\ servicio\ en\ (t, t+\Delta t])$$

$$+ P(n+1\ clientes\ en\ t, no\ llegadas + un\ servicio\ en\ (t, t+\Delta t])$$

$$+ P(n-1\ clientes\ en\ t, 1\ llegada + no\ servicios\ en\ ]) + o(\Delta t)$$

Puesto que los tiempos de llegada y servicio son independientes entre sí y a su vez del estado del sistema t (propiedad de incrementos independientes), se tiene que, para cada $n \geq 1$

$$p_n(t+\Delta t) = p_n(t)P(no\ llegadas\ en\ (t, t+\Delta t])P(no\ servicios\ en\ (t, t+\Delta t)$$

$$+ p_n(t)P(1\ llegada\ en\ (t, t+\Delta t])P(1\ servicio\ en\ (t, t+\Delta t])$$

$$+ p_{n+1}(t)P(no\ llegadas\ en\ (t, t+\Delta t])P(1\ servicio\ en\ (t, t+\Delta t])$$

$$+ p_{n-1}(t)P(1\ llegada\ en\ (t, t+\Delta t])P(no\ servicios\ en\ (t, t+\Delta t]) + o(\Delta t)$$

Se puede suponer que los tiempos entre llegadas son exponenciales de parámetro λ y los tiempos de servicio son exponenciales de parámetro μ [9]. Utilizando las propiedades de los correspondientes Procesos de Poisson, podemos escribir:

$$p_n(t+\Delta t) = p_n(t)\big(1-\lambda\Delta t + o(\Delta t)\big)\big(1-\mu\Delta t + o(\Delta t)\big)$$

$$+ p_n(t)\big(\lambda\Delta t + o(\Delta t)\big)\big(\mu\Delta t + o(\Delta t)\big)$$

$$+ p_{n+1}(t)(1-\lambda\Delta t + o(\Delta t))(\mu\Delta t + o(\Delta t)))$$

$$+ p_{n-1}(t)\big(\lambda\Delta t + o(\Delta t)\big)\big(1-\mu\Delta t + o(\Delta t)\big) + o(\Delta t)$$

Uniendo todos los términos $o(\Delta t)$ y teniendo en la cuenta que los términos $(\Delta t)^2$ son también $o(\Delta t)$, se tiene:

$$p_n(t+\Delta t) = p_n(t)(1-\lambda\Delta t - \mu\Delta t) + p_{n+1}(t)(\mu\Delta t) + p_{n-1}(\lambda\Delta t) + o(\Delta t) \quad (6)$$

Para el caso $n = 0$ el análisis es similar teniendo en la cuenta que el caso relativo a $p_{n-1}(t)$ no se puede dar, obteniéndose en este caso la siguiente ecuación:

$$p_0(t+\Delta t) = p_0(t)(1-\lambda\Delta t) + p_1(t)(\mu\Delta t) + o(\Delta t) \quad (7)$$

Las ecuaciones (6) y (7) constituyen el sistema de ecuaciones en diferencia para el caso $M/M/1$. Estas ecuaciones en diferencia lo son tanto con respecto a t como a n.

- **Ecuaciones diferenciales en diferencia**

El sistema de ecuaciones formado por (6) y (7) se puede reescribir del siguiente modo:

$$\begin{cases} p_n(t+\Delta t) - p_n(t) = -(\lambda+\mu)p_n(t)\Delta t + \mu p_{n+1}(t)\Delta t + \lambda p_{n-1}(t)\Delta t + o(\Delta t) & n \geq 1 \\ p_0(t+\Delta t) - p_0(t) = -\lambda p_0(t)\Delta t + \mu p_1(t)\Delta t + o(\Delta t) \end{cases}$$

Si se divide por Δt y se toma límites cuando $\Delta t \longrightarrow 0$, se tiene que:

$$\begin{cases} \dfrac{dp_n(t)}{dt} = -(\lambda+\mu)p_n(t) + \mu p_{n+1}(t) + \lambda p_{n-1}(t), & n \geq 1 \\ \dfrac{dp_0(t)}{dt} = -\lambda p_0(t) + \mu p_1(t) \end{cases} \tag{8}$$

- **Obtención de las probabilidades estacionarias p_n**

Se supone que el sistema alcanza un estado estacionario. Esto significa que cuando $t \longrightarrow \infty$ la probabilidad $p_n(t)$ se vuelve independiente del tiempo. Por lo tanto, $\frac{dp_n(t)}{dt}$ tendería a cero, y el sistema de ecuaciones (9) quedaría como el siguiente sistema de ecuaciones en diferencia:

$$\begin{cases} 0 = -(\lambda+\mu)p_n + \mu p_{n+1} + \lambda p_{n-1}, & n \geq 1 \\ 0 = -\lambda p_0 + \mu p_1 \end{cases}$$

Escrito de otro modo,

$$\begin{cases} p_{n+1} = \dfrac{\lambda+\mu}{\mu}p_n - \dfrac{\lambda}{\mu}p_{n-1}, & n \geq 1 \\ p_1 = \dfrac{\lambda}{\mu}p_0 \end{cases} \tag{9}$$

Este sistema de ecuaciones en diferencia en una variable (n), se puede resolver utilizando un procedimiento iterativo [9], llegando a:

$$p_2 = \frac{\lambda+\mu}{\mu}p_1 - \frac{\lambda}{\mu}p_0 = \frac{\lambda+\mu}{\mu}\left(\frac{\lambda}{\mu}p_0\right) - \frac{\lambda}{\mu}p_0$$

$$p_2 = \left(\frac{\lambda}{\mu}\right)^2 p_0$$

Escrito de manera general sería:

$$p_n = \left(\frac{\lambda}{\mu}\right)^n p_0 \tag{10}$$

Para completar el análisis falta obtener el valor de p_0. Para ellos se utiliza que $\{p_n\}\ n \geq 0$ debe ser una función puntual de probabilidad, por lo tanto,

$$\sum_{n\geq 0} p_n = 1$$

Utilizando (10) se tiene que

$$\sum_{n\geq 0}\left(\frac{\lambda}{\mu}\right)^n p_0 = 1,$$

De donde se sigue que

$$\frac{1}{p_0} = \sum_{n\geq 0}\left(\frac{\lambda}{\mu}\right)^n$$

La expresión $\sum_{n\geq 0}\left(\frac{\lambda}{\mu}\right)^n$ corresponde a una serie geométrica que converge si $\left|\frac{\lambda}{\mu}\right| = \frac{\lambda}{\mu} < 1$, o lo que es lo mismo, que $\lambda > \mu$ significa que el tiempo medio de llegada es mayor que el tiempo medio de servicio, y por lo tanto el estado del sistema crecerá indefinidamente. Entonces, no existe el estado estacionario si $\lambda = \mu$, con posible explicación relacionada a que en este caso conforme la cola crece se le hace más difícil al servidor disminuir el tamaño de la misma, ya que la tasa de servicio es menor que la de llegadas [9].

En definitiva, si $\rho = \frac{\lambda}{\mu} < 1$ se tiene que:

$$\frac{1}{p_0} = \sum_{n\geq 0}\left(\frac{\lambda}{\mu}\right)^n = \sum_{n\geq 0}\rho^n = \frac{1}{1-\rho}$$

De donde se sigue que

$$p_0 = 1 - \rho$$

De acuerdo con lo anterior, la solución general para el estado estacionario sería:

$$p_n = \rho^n(1-\rho) \tag{11}$$

- **Medidas de eficiencia**

A través de la distribución de probabilidad del estado del sistema para el caso estacionario se puede obtener algunas medidas para calibrar la eficiencia del sistema, como son el número medio de clientes en el sistema y el número medio de clientes en cola (siempre suponiendo que el sistema se encuentra en estado estacionario) [9].

Sea $\mathcal{L}$ la variable aleatoria "número de clientes en el sistema" y su esperanza matemática $L = E(\mathcal{L})$:

$$L = \sum_{n\geq 0} np_n = \sum_{n\geq 0} n(1-\rho)\rho^n = \sum_{n\geq 1} n(1-\rho)\rho^n = \rho \sum_{n} n(1-\rho)\rho^{n-1}$$

$$L = \rho \frac{1}{(1-\rho)}$$

O de forma equivalente:

$$L = \frac{\lambda}{\mu - \lambda}$$

Por otro lado, sea $\mathcal{L}_q$ la variable aleatoria "número de clientes en la cola" y L_q su esperanza matemática que se calcula $L_q = E(\mathcal{L})_q$. Se debe observar que

$$\mathcal{L}_q = \begin{cases} 0, & si\ \mathcal{L} = 0 \\ \mathcal{L} - 1, & si\ \mathcal{L} \geq 1 \end{cases}$$

Por lo tanto, se tiene que

$$L_q = 0p_0 + \sum_{n\geq 1}(n-1)p_n = \sum_{n\geq 1} np_n - \sum_{n\geq 1} p_n = L - (1-p_0) = \frac{\rho}{1-\rho} - \rho$$

O lo que es lo mismo,

$$L_q = \frac{\lambda^2}{\mu(\mu-\lambda)} \tag{12}$$

- **Distribución de los tiempos de espera**

Se obtiene información del tiempo promedio que debe esperar un cliente en el sistema y en la cola para ser servido [9].

Se denota a $\mathcal{W}_q$ a la variable aleatoria que representa el tiempo que pasa un cliente en cola, y sea W_q su esperanza.

Se calcula el tiempo medio que espera un individuo $W_q = E(\mathcal{W}_q)$ condicionado por $\mathcal{L}$, el número de clientes en el sistema cuando llega al mismo.

$$W_q = E(\mathcal{W}_q) = E(E(\mathcal{W}_q|\mathcal{L}))$$

$$= \sum_{n=0}^{\infty} E(\mathcal{W}_q|\mathcal{L} = n)P(\mathcal{L} = n) \tag{13}$$

Se obtiene el valor de $E(\mathcal{W}_q|\mathcal{L} = n)$. Si no hay clientes en el sistema $(n = 0)$, claramente el tiempo de espera en cola es 0. Si hay $n \geq 1$ clientes en el sistema, entonces el nuevo cliente tiene que esperar a que se completen los n servicios que tiene delante; el de los $n - 1$ clientes que hay en cola delante de él, más el del cliente que se sirve Los $n - 1$ clientes que están en cola tardarán cada uno un tiempo exponencial de parámetro μ, para el cliente que está siendo servido, como la propiedad exponencial tiene la propiedad de pérdida de memoria, el tiempo que le queda una vez que se ha producido la llegada del nuevo cliente sigue siendo exponencial de parámetro μ [9]. Por lo tanto, cuando el nuevo cliente llega, tiene delante n clientes con tiempo exponenciales de parámetro μ, por lo que,

$$E(\mathcal{W}_q|\mathcal{L} = n) = \frac{n}{\mu}$$

Volviendo de nuevo a la ecuación (13),

$$W_q = \sum_{n=0}^{\infty} E(\mathcal{W}_q|\mathcal{L} = n)\, P(\mathcal{L} = n) = \sum_{n=1}^{\infty} \frac{n}{\mu} p_n = \sum_{n=1}^{\infty} \frac{n}{\mu} \rho^n (1 - \rho)$$

$$= \frac{\rho}{\mu} \sum_{n=1}^{\infty} \rho^{n-1}(1 - \rho) = \frac{1}{\mu} \frac{\rho}{1 - \rho}$$

$$W_q = \frac{\lambda}{\mu(\mu - \lambda)}$$

Aunque usualmente la característica de interés es la esperanza del tiempo de espera en cola, para el cálculo de otras medidas se puede comprobar que la función de distribución de la variable $\mathcal{W}_q$ viene dada por:

$$F\mathcal{W}_q(t) = \begin{cases} 1-\rho & si\ t = 0 \\ 1-\rho e^{-\mu(1-\rho)t}, & si\ t > 0 \end{cases}$$

Considerando a $\mathcal{W}$ como la variable aleatoria continua que representa el tiempo que pasa un cliente en el sistema y sea W su esperanza.

$$W = E(\mathcal{W}) = E(E(\mathcal{W}|\mathcal{L}))$$

$$= \sum_{n=0}^{\infty} E(\mathcal{W}|\mathcal{L} = n)\, P(\mathcal{L} = n) \qquad (14)$$

Análogamente al caso anterior, se demuestra que

$$E(\mathcal{W}|\mathcal{L} = n) = \frac{n+1}{\mu}$$

Desde la ecuación (14),

$$W = \sum_{n=0}^{\infty} E(\mathcal{W}|\mathcal{L} = n)\, P(\mathcal{L} = n) = \sum_{n=0}^{\infty} \frac{n+1}{\mu} p_n$$

$$= \frac{1}{\mu}\left(\sum_{n=0}^{\infty} np^n(1-\rho) + \sum_{n=0}^{\infty} p^n(1-\rho)\right)$$

$$= \frac{1}{\mu}\left(\frac{\rho}{1-\rho} + 1\right) = \frac{1}{\mu}\frac{1}{1-\rho} = \frac{1}{\mu-\lambda}$$

La función de densidad de la variable $\mathcal{W}$ está dada por:

$$F_{\mathcal{W}}(t) = (\mu-\lambda)e^{-(\mu-\lambda)t}, \qquad t > 0$$

5.6.2. Teorema de Little

Resultado fundamental de la teoría de colas, que establece una relación entra la tasa efectiva de entrada de usuarios en un sistema, el número medio de usuarios

en dicho sistema y el tiempo medio total que pasan en el mismo [10]. Parte de su importancia radica, además, en que se puede aplicar a casi cualquier sistema.

Sea el caso de un sistema como el que se muestra en la Figura 5, en donde la gráfica de arriba se refiere a la llegada de los clientes y la de abajo representa las salidas. De allí cabe notar que la diferencia horizontal brinda los tiempos tanto en cola como en el sistema, siendo éste la mayor distancia horizontal. De la diferencia vertical se puede extraer el tamaño de cola y el tamaño del sistema que también viene siendo la mayor distancia en vertical.

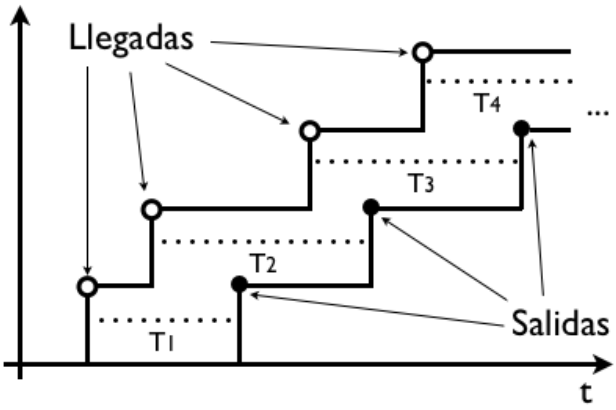


Figura 5. Ilustración del Teorema de Little *[10]*.

- $\boldsymbol{L = \lambda W}$**:** esta expresión se conoce comúnmente como la fórmula de Little.

- $\boldsymbol{W = W_q + \frac{1}{\mu}}$**:** si Y es la variable aleatoria que representa el tiempo de servicio, entonces $\mathcal{W} = \mathcal{W}_q + Y$. Puesto que la esperanza de la suma es la suma de las esperanzas, se tiene que $E(\mathcal{W}) = E(\mathcal{W}_q) + E(Y)$, obteniéndose el resultado.

- $\boldsymbol{L_q = \lambda W_q}$ **:** se supone que un cliente llega al sistema. En promedio entrará al servicio después de un tiempo W_q. Se supone que justo cuando va a entrar al servicio se da la vuelta y cuenta los clientes que están en cola detrás de él; en promedio ese número sería L_q. Puesto que en promedio cada uno de los L_q que están en la cola han tardado en llegar $\frac{1}{\lambda}$ respecto del anterior, el tiempo que ha estado esperando el cliente en cuestión en cola es $L_q\left(\frac{1}{\lambda}\right)$.

- $\boldsymbol{L = L_q + \frac{\lambda}{\mu}}$**:** es consecuencia inmediata de las anteriores.

- $\boldsymbol{W_q = \frac{1}{\mu} L}$**:** justo cuando llega un cliente al sistema espera encontrarse L clientes delante de él. Para empezar su servicio tendrá que esperar a que finalice el

servicio de los L anteriores. Puesto que el tiempo de servicio promedio es $\frac{1}{\mu}$, el tiempo medio que espera en la cola es $\frac{1}{\mu}L$ (observar que se está utilizando la propiedad de pérdida de memoria de la función exponencial para asegurar que un cliente que está siendo servido cuando el cliente en cuestión se incorpora a la cola también tardará en finalizar su servicio un tiempo exponencial μ).

5.6.3. Modelo $M/M/1/K$

A continuación, se va a analizar una modificación del modelo $M/M/1$ que se basa en suponer que la capacidad del sistema está limitada a K clientes [9].

Las ecuaciones de estado del sistema, $p_n(t)$, dadas para el modelo $M/M/1$, siguen siendo válidas si $n < K$. Ahora se analiza el caso $n = K$. Para que el sistema esté en el estado K en el instante $t + \Delta t$ pueden darse tres situaciones [9]:

- Que el sistema esté en el estado K en el instante t, y en el intervalo $(t, t + \Delta t]$ no se produzcan llegadas ni servicios. La probabilidad asociada a esta situación es:

$$p_K(t)(1 - \lambda\Delta t + o(\Delta t))(1 - \mu\Delta t + o(\Delta t))$$

- Que el sistema esté en estado K en el instante t, y en el intervalo $(t, t + \Delta t]$ se produzca una llegada y ningún servicio. Como la capacidad máxima del sistema es K, el cliente que llega es rechazado y el sistema permanece en estado K. La probabilidad asociada a esta situación es

$$p_K(t)(\lambda\Delta t + o(\Delta t))(1 - \mu\Delta t + o(\Delta t))$$

- Que el sistema esté en el estado $k - 1$ en el instante t, y en el intervalo $(t, t + \Delta t]$ se produzca una llegada y ningún servicio. La probabilidad asociada es

$$p_{K-1}(t)(\lambda\Delta t + o(\Delta t))(1 - \mu\Delta t + o(\Delta t))$$

Por lo tanto,

$$\begin{aligned} p_K(t + \Delta t) \quad &= p_K(t)\big(1 - \lambda\Delta t + o(\Delta t)\big)\big(1 - \mu\Delta t + o(\Delta t)\big) \\ &+ p_K(t)\big(\lambda\Delta t + o(\Delta t)\big)\big(1 - \mu\Delta t + o(\Delta t)\big) \\ &+ p_{K-1}(t)\big(\lambda\Delta t + o(\Delta t)\big)\big(1 - \mu\Delta t + o(\Delta t)\big) \\ &= p_K(t)(1 - \mu\Delta t) + p_{K-1}(t)(\lambda\Delta t))(1 - \mu\Delta t) + o(\Delta t) \end{aligned}$$

De lo anterior se obtiene la ecuación diferencial

$$\frac{dp_K(t)}{dt} = -\mu p_K(t) + \lambda p_{K-1}(t)$$

Y tomando el límite cuando $t \longrightarrow \infty$, se obtiene para el caso estacionario

$$p_K = \frac{\lambda}{\mu} p_{k-1}$$

El sistema de ecuaciones en diferencia para las probabilidades de estado en situación estacionaria del modelo $M/M/1/K$ es:

$$\begin{cases} p_1 = \dfrac{\lambda}{\mu} p_0 \\ p_{n+1} = \dfrac{\lambda+\mu}{\mu} p_n - \dfrac{\lambda}{\mu} p_{n-1}, \qquad 1 \leq n \leq K-1 \\ p_K = \dfrac{\lambda}{\mu} p_{K-1} \end{cases}$$

Utilizando el mismo método de solución para el modelo $M/M/1$, se sabe que

$$p_n = \left(\frac{\lambda}{\mu}\right)^n p_0, \quad 0 \leq n \leq K-1$$

De la expresión de p_K se deduce que la anterior relación también se verifica para $n = K$. Por lo tanto,

$$p_n = (\rho)^n p_0 \qquad 0 \leq n \leq K,$$

Siendo $\rho = \frac{\lambda}{\mu}$. El valor de p_0 se puede obtener de la condición $\sum_{n=0}^{K} \rho^n p_0 = 1$, de donde se sigue que

$$p_0 = \frac{1}{\sum_{n=0}^{K} \rho^n}$$

El denominador de la anterior expresión corresponde a la de una serie geométrica finita cuyo valor es

$$\sum_{n=0}^{K} \rho^n = \begin{cases} \dfrac{1-\rho^{K+1}}{1-\rho}, & \rho \neq 1 \\ K+1 & \rho = 1 \end{cases}$$

Por lo tanto,

$$p_0 = \begin{cases} \dfrac{1-\rho}{1-\rho^{K+1}}, & \rho \neq 1 \\ \dfrac{1}{K+1} & \rho = 1 \end{cases}$$

De donde se sigue que la expresión final de las probabilidades de estado es:

$$p_n = \begin{cases} \dfrac{(1-\rho)\rho^n}{1-\rho^{K+1}}, & \rho \neq 1 \\ \dfrac{1}{K+1} & \rho = 1 \end{cases} \tag{15}$$

Se puede observar que en este caso, la solución para el estado estacionario existe incluso si $\rho \geq 1$. Intuitivamente esto se debe a que la limitación en la capacidad del sistema provoca que éste no se desborde [9].

- **Medidas de eficiencia**

Primero que todo, se obtiene el número medio de clientes en el sistema L, a través de la expresión $L = \sum_{n=0}^{K} np_n$. Se observa que si $\rho = 1$, entonces de la fórmula (15) se sigue:

$$L = \sum_{n=0}^{K} np_n = \sum_{n=0}^{K} n\frac{1}{K+1} = \frac{1}{K+1}\left(\sum_{n=0}^{K} n\right) = \frac{1}{K+1}\frac{K(k+1)}{2} = \frac{K}{2}$$

Si $\rho \neq 1$, entonces

$$\begin{aligned} L = \sum_{n=0}^{K} np_n &= \sum_{n=0}^{K} np_0p^n = p_0\rho\sum_{n=1}^{K} n\rho^{n-1} = p_0\rho\sum_{n=0}^{K} n\rho^{n-1} \\ &= p_0\rho\frac{d}{dp}\left(\sum_{n=0}^{K}\rho^n\right) = p_0\rho\frac{d}{dp}\left(\frac{1-\rho^{K+1}}{1-\rho}\right) \\ &= p_0\rho\frac{1-(K+1)\rho^K+K\rho^{K+1}}{(1-\rho)^2} \\ &= \frac{\rho[1-(K+1)\rho^K+K\rho^{K+1}]}{(1-\rho^{K+1})(1-\rho)} \end{aligned}$$

Para el tamaño medio de la cola L_q, se obtiene

$$L_q = 0p_0 + \sum_{n=1}^{K}(n-1)p_n = \sum_{n=1}^{K} np_n + \sum_{n=1}^{K} p_n = \sum_{n=0}^{K} np_n + (1-p_0) = L - (1-p_0)$$

De donde se sigue que:

$$L_q = L - (1 - p_o) = \begin{cases} L - \dfrac{\rho(1-\rho^K)}{1-\rho^{K+1}}, & \rho \neq 1 \\ \dfrac{K(K-1)}{2(K+1)}, & \rho = 1 \end{cases}$$

Para el tiempo medio de estancia en el sistema de un cliente W, condicionado por el número de clientes en el sistema cuando el cliente se incorpora al mismo, se observa que para que el cliente no sea rechazado tiene que haber a lo sumo $K-1$ clientes en el sistema, luego, la variable por la que hay que condicionar no es $\mathcal{L}$ sino $\overline{\mathcal{L}} = (\mathcal{L}|\mathcal{L} \leq K-1)$, cuyas probabilidades puntuales son:

$$\overline{p_n} = P(\mathcal{L} = n|\mathcal{L} \leq K-1) = \frac{P(\mathcal{L}=n)}{P(\mathcal{L} \geq K-1)} = \frac{p_n}{1-p_K}, \qquad n \leq K-1$$

$$W = \sum_{n=0}^{K-1} E(W|\overline{\mathcal{L}} = n)\overline{p_n}$$

Se puede observar que para n fijo, se tiene que $E\left(W\middle|\overline{\mathcal{L}} = n\right) = \frac{n+1}{\mu}$, luego

$$\begin{aligned} W &= \sum_{n=0}^{K-1} E\left(W\middle|\overline{\mathcal{L}} = n+1\right)\overline{p_n} = \sum_{n=0}^{K-1} \frac{n+1}{\mu}\frac{p_n}{1-p_K} \\ &= \frac{1}{\mu(1-p_K)}\left(\sum_{n=0}^{K-1}(n+1)p_n\right) = \frac{1}{\mu(1-p_K)}\left(\sum_{n=0}^{K-1} np_n + \sum_{n=0}^{k-1} p_n\right) \\ &= \frac{1}{\mu(1-p_K)}([L - Kp_K] + [1-p_K]) \end{aligned}$$

Se puede comprobar que $(L - Kp_K) + (1-p_K) = L + 1 - (K+1)p_K = \frac{L}{\rho}$, luego

$$W = \frac{1}{\mu(1-p_K)}([L-Kp_K] + [1-p_K]) = \frac{1}{\mu(1-p_K)}\frac{L}{\rho} = \frac{L}{\lambda(1-p_K)} \qquad (16)$$

Para el cálculo del tiempo medio de espera en cola W_q, se puede utilizar la relación $W = W_q + \frac{1}{\mu}$. Así pues, utilizando la ecuación (16) se puede comprobar que:

$$W_q = \frac{L_q}{\lambda(1-p_K)}$$

5.7. MODELOS DE COLAS EXPONENCIALES CON VARIOS SERVIDORES EN PARALELO

En este numeral se analizan modelos de colas en los cuales las llegadas se producen de acuerdo con un proceso de Poisson de parámetro λ, hay c servidores que atienden a los clientes cuyos tiempos de servicio son independientes e idénticamente distribuidos, exponenciales de parámetro μ. Estas colas se ajustan a un modelo de nacimiento y muerte en el que $\lambda_n = \lambda$, para todo n. Por otro lado, μ_n es el número de servicios que finalizan por unidad de tiempo cuando en el sistema hay n clientes. Si $n > c$, entonces los c servidores están ocupados y como cada uno de ellos sirve a μ clientes por unidad de tiempo, en total se finalizan $c\mu$ servicios por unidad de tiempo. Si $1 \leq n \leq c$, entonces sólo n de los c canales están ocupados, y por lo tanto, la tasa de finalización de servicios es $n\mu$ [9]. En definitiva,

$$\mu_n = \begin{cases} n\mu, & 1 \leq n \leq c \\ c\mu, & n > c \end{cases}$$

$$p_n = \begin{cases} \dfrac{\lambda^n}{n!\,\mu^n} p_0, & 1 \leq n \leq c \\ \dfrac{\lambda^n}{c!\,c^{n-c}\mu^n} p_0, & n \geq c \end{cases}$$

Utilizando la condición de que $\sum_{n=0}^{\infty} p_n = 1$, se puede obtener el valor de p_0.

$$p_0 \left(\sum_{n=0}^{c-1} \frac{\lambda^n}{n!\,\mu^n} + \sum_{n=c}^{\infty} \frac{\lambda^n}{c!\,c^{n-c}\mu^n} \right) = 1$$

Denotando $r = \frac{\lambda}{\mu}$,$\rho = \frac{\lambda}{c\mu} = \frac{r}{c}$, se tiene que la expresión anterior queda

$$p_0 \left(\sum_{n=0}^{c-1} \frac{r^n}{n!} + \sum_{n=c}^{\infty} \frac{r^n}{c!\,c^{n-c}} \right) = 1$$

Ahora, para la serie $\sum_{n=c}^{\infty} \frac{r^n}{c!c^{n-c}}$ se tiene:

$$\sum_{n=c}^{\infty} \frac{r^n}{c!\,c^{n-c}} = \frac{r^c}{c!} \sum_{n=c}^{\infty} \left(\frac{r}{c}\right)^{n-c} = \frac{r^c}{c!} \sum_{m=0}^{\infty} \rho^m \qquad (17)$$

Que converge a $\frac{r^c}{c!}\frac{1}{1-\rho}$ siempre que $\rho = \frac{r}{c} < 1$.

Por lo tanto, para el modelo $M/M/c$, si se verifica que $\rho = \frac{r}{c} = \frac{\lambda}{c\mu} < 1$, entonces la solución estacionaria existe y el valor de p_0 está dado por

$$p_0 = \left(\sum_{n=0}^{c-1} \frac{r^n}{n!} + \frac{cr^c}{c!\,(c-r)} \right)^{-1} = \left(\sum_{n=0}^{c-1} \frac{1}{n!} \left(\frac{\lambda}{\mu}\right)^n + \frac{1}{c!} \left(\frac{\lambda}{\mu}\right)^c \left(\frac{c\mu}{c\mu - \lambda} \right) \right)^{-1} \qquad (18)$$

- **Medidas de eficiencia**

Tamaño esperado de la cola, L_q

El tamaño de la cola $\mathcal{L}_q$ es cero si el número de clientes en el sistema, n, es menor o igual que c y $n - c$ en caso contrario. Por lo tanto,

$$L_q = E(\mathcal{L}_{\mathrm{II}}) = \sum_{n=c}^{\infty} (n-c) p_n = \sum_{n=c}^{\infty} \frac{n}{c^{n-c} c!} r^n p_0 - \sum_{n=c}^{\infty} \frac{c}{c^{n-c} c!} r^n p_0 \qquad (19)$$

Examinando la primera serie del lado derecho de (19)

$$\frac{p_0}{c!} \sum_{n=c}^{\infty} \frac{n}{c^{n-c}} r^n = \frac{p_0}{c!} \frac{r^{c+1}}{c} \left[\sum_{n=c}^{\infty} (n-c) \left(\frac{r}{c}\right)^{n-c-1} + \sum_{n=c}^{\infty} c \left(\frac{r}{c}\right)^{n-c-1} \right]$$

$$= \frac{p_0}{c!} \frac{r^{c+1}}{c} \left[\sum_{n=0}^{\infty} n \left(\frac{r}{c}\right)^{n-1} + \sum_{n=0}^{\infty} c \left(\frac{r}{c}\right)^{n-1} \right]$$

$$= \frac{p_0}{c!} \frac{r^{c+1}}{c} \left[\sum_{n=0}^{\infty} n \left(\frac{r}{c}\right)^{n-1} + c \frac{c}{r} \sum_{n=0}^{\infty} c \left(\frac{r}{c}\right)^{n} \right]$$

$$= \frac{p_0}{c!} \frac{r^{c+1}}{c} \left[\frac{1}{\left(1 - \frac{r}{c}\right)^2} + \frac{\frac{c^2}{r}}{1 - \frac{r}{c}} \right]$$

El valor de la segunda serie del lado derecho de (19), se obtiene a partir (17)

$$\sum_{n=c}^{\infty} \frac{c}{c^{n-c} c!} r^n p_o = p_0 c \sum_{n=c}^{\infty} \frac{r^n}{c^{n-c} c!} = \frac{p_0 c r^c}{c! \left(1 - \frac{r}{c}\right)}$$

Reemplazando en la ecuación (19), queda

$$L_q = p_0 \frac{r^{c+1}}{c \cdot c!} \left[\frac{1}{\left(1 - \frac{r}{c}\right)^2} + \frac{\frac{c^2}{r}}{1 - \frac{r}{c}} + \frac{\frac{c^2}{r}}{1 - \frac{r}{c}} \right] = p_0 \frac{r^{c+1}}{c \cdot c!} \left[\frac{1}{\left(1 - \frac{r}{c}\right)^2} \right]$$

Finalmente,

$$L_q = \left[\frac{\left(\frac{\lambda}{\mu}\right)^c \lambda\mu}{(c-1)!\,(c\mu - \lambda)^2} \right] p_0 \tag{20}$$

Tiempo medio de espera en la cola

Asumiendo de igual manera como para el modelo $M/M/1$ que la disciplina de la cola es FIFO y haciendo uso de la fórmula de Little, se obtiene:

$$W_q = \frac{L_q}{\lambda} = \left[\frac{\left(\frac{\lambda}{\mu}\right)^c \mu}{(c-1)!\,(c\mu - \lambda)^2} \right] p_0 \tag{21}$$

Tiempo medio de estancia en el sistema

Utilizando la relación general entre W y W_q y el resultado obtenido en (21), se obtiene lo siguiente:

$$W = W_q + \frac{1}{\mu} = \left[\frac{\left(\frac{\lambda}{\mu}\right)^c \mu}{(c-1)!\,(c\mu - \lambda)^2} \right] p_0 + \frac{1}{\mu} \tag{22}$$

Número medio de clientes en el sistema

Usando la fórmula de Little y la ecuación anterior, se obtiene lo siguiente:

$$L = \lambda W = \left[\frac{\left(\frac{\lambda}{\mu}\right)^c \lambda\mu}{(c-1)!\,(c\mu - \lambda)^2} \right] p_0 + \frac{\lambda}{\mu} \tag{23}$$

5.7.1. Modelo $M/M/c$ con fuente de entrada finita

Se modifica el modelo $M/M/c$ de tal forma que su fuente de entrada sea limitada, es decir, el tamaño de la población de potenciales clientes es finito. Tomando a N como el tamaño de dicha población. De este modo, cuando en el sistema se encuentran n clientes, restan solo $N-n$ clientes potenciales en la fuente de entrada [9].

En el modelo con población finita los clientes alternan entre estar dentro y fuera del sistema, así pues, por analogía con el modelo $M/M/c$ se supone que el tiempo que pasa cada miembro fuera del sistema es una variable exponencial de parámetro λ. Cuando n miembros están dentro, $N-n$ están fuera, por lo tanto la distribución de probabilidad del tiempo que falta para la próxima llegada al sistema es el mínimo de $N-n$ variables exponenciales independientes de parámetro λ. Se puede demostrar que esta distribución se ajusta a una exponencial de parámetro $\lambda(N-n)$ [9].

$$\lambda_n = \begin{cases} (N-n)\lambda, & 0 \le n \le N \\ 0, & n > N \end{cases}$$

$$\mu_n = \begin{cases} n\mu\lambda, & 1 \le n \le c \\ c\mu\lambda, & c \le n \le N \\ 0, & n > N \end{cases}$$

- **Medidas de eficiencia para el caso $M/M/c$ con población finita**

Se obtienen las medidas que miden la eficiencia dentro del sistema como lo son el tiempo total en el sistema, tiempo en la cola, tamaño del sistema y tamaño de la cola

$$p_0 = \left[\sum_{n=0}^{c-1} \binom{N}{n} \left(\frac{\lambda}{\mu}\right)^n + \sum_{n=c}^{N} \binom{N}{n} \frac{n!}{c!\, c^{n-c}} \left(\frac{\lambda}{\mu}\right)^n\right]^{-1}$$

$$p_n = \begin{cases} \binom{N}{n} \left(\frac{\lambda}{\mu}\right)^n p_0, & 0 \le n \le c \\ \binom{N}{n} \frac{n!}{c^{n-c} c!} \left(\frac{\lambda}{\mu}\right)^n p_0, & c \le n \le N \end{cases}$$

$$L = p_0 \left[\sum_{n=0}^{c-1} n \binom{N}{n} \left(\frac{\lambda}{\mu}\right)^n + \frac{1}{c!} \sum_{n=c}^{N} n \binom{N}{n} \frac{n!}{c^{n-c}} \left(\frac{\lambda}{\mu}\right)^n\right]$$

$$L_q = L - c + p_0 \sum_{n=0}^{c-1} (c-n) \binom{N}{n} \left(\frac{\lambda}{\mu}\right)^n$$

Usando las siguientes ecuaciones, se encuentra a W_q y W, donde $\bar{\lambda} = \lambda(N - L)$:

$$W_q = \frac{L_q}{\bar{\lambda}} = \frac{L - c + p_0 \sum_{n=0}^{c-1}(c-n)\binom{N}{n}\left(\frac{\lambda}{\mu}\right)^n}{\lambda(N-L)}$$

$$W = \frac{L}{\bar{\lambda}} = \frac{p_0\left[\sum_{n=0}^{c-1} n\binom{N}{n}\left(\frac{\lambda}{\mu}\right)^n + \frac{1}{c!}\sum_{n=c}^{N} n\binom{N}{n}\frac{n!}{c^{n-c}}\left(\frac{\lambda}{\mu}\right)^n\right]}{\lambda(N-L)}$$

5.7.2. Modelo $M/G/1$

En este modelo, el sistema de colas tiene un servidor, las llegadas se producen según un proceso de Poisson de tasa λ y los clientes tienen tiempos de servicio independientes e idénticamente distribuidos de media $\frac{1}{\mu}$ y varianza σ^2.

Cualquier sistema de colas de este tipo alcanza en algún momento el estado estable si $\rho = \frac{\lambda}{\mu} < 1$.

$$p_0 = 1 - \rho$$

$$L_q = \frac{\lambda^2\sigma^2 + \rho^2}{2(1-\rho)}$$

$$L = \frac{\lambda^2\sigma^2 + \rho^2}{2(1-\rho)} + \rho$$

$$W_q = \frac{\lambda^2\sigma^2 + \rho^2}{2\lambda(1-\rho)}$$

$$W = \frac{\lambda^2\sigma^2 + \rho^2}{2\lambda(1-\rho)} + \frac{1}{\mu}$$

Para este sistema el funcionamiento del servidor tiene una gran trascendencia en la eficiencia global del sistema.

CAPÍTULO 6.

REDES DE PETRI

El concepto de las Redes de Petri (RdP) fue introducido por el Dr. Carl Adam Petri en el año de 1962 para su disertación doctoral "*kommunication mit automaten"* (Comunicación con autómatas). Las RdP son un formalismo para modelar, analizar, simular, controlar y evaluar el comportamiento de sistemas concurrentes, paralelos, no determinísticos, secuenciales, de eventos discretos, estocásticos, entre otros. Las Redes de Petri combinan una sólida base matemática con una representación gráfica que simplifica su entendimiento [17].

Las RdP se componen por los siguientes elementos [18]:

- **Transiciones (T):** representan los procesos del sistema.
- **Lugares (P):** representan las condiciones necesarias para que un proceso se ejecute.
- **Arcos:** relacionan las condiciones con procesos y se les asocia de forma individual un peso.
- **Marcas:** Si se encuentran presentes en un lugar, indican que se verifica la condición que representa dicho lugar.

Una RdP se describe como una quíntupla $PN = \{P, T, F, W, M_0\}$ donde:

- $P = \{p_1, p_2, p_i, \dots, p_m\}$ es el conjunto finito de lugares de la red.
- $T = \{t_1, t_2, t_j, \dots, t_n\}$ es el conjunto finito de transiciones de la red.
- $F \subseteq (P \times T) \cup (T \times P)$ es el conjunto de arcos que definen el flujo de la red.
- $W: F \longrightarrow \{1,2,3, \dots\}$ es la función de peso.
- $M_0: P \rightarrow \{0,1,2,3, \dots\}$ es el marcado inicial de la red.

El comportamiento de un sistema está representado en todo instante por el marcado actual de la red, siendo este marcado indicativo del estado y los cambios que se presentan en el sistema La evolución en el marcado se rige por las siguientes reglas de transición:

1. Una transición está habilitada si en cada lugar de entrada a ella, hay al menos un número de marcas igual al peso de los arcos que lo conectan con dicha transición.

2. Una transición sensibilizada puede ser disparada dependiendo de si su evento asociado ocurre.

3. El disparo de una transición sensibilizada remueve marcas de cada lugar de entrada a la transición y adiciona marcas a cada lugar de salida de la transición, donde es el peso del arco de salida que une a t_j con p_i.

La Figura 6 es un ejemplo de una RdP y en ella se puede observar los elementos que la componen.

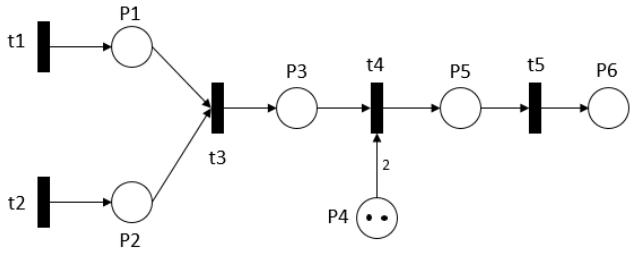


Figura 6. Elementos que componen una RdP *[19]*.

En la figura anterior, se puede notar que el peso del arco que conecta el lugar P4 con t4 tiene un valor de 2, pero el peso en los demás arcos es de 1 (Por lo tanto, se omite poner este valor). En esta red sólo el lugar P4 tiene marcas al inicio por lo que el vector de marcado inicial en este caso es: $M_0 = \{0,0,0,2,0,0\}$ el cual denota claramente la existencia de 2 marcas en el lugar 4 y ninguna para los demás.

6.1. TIPOS DE TRANSICIONES Y LUGARES

Cuando una transición no posee ningún lugar de entrada se dice que es una *Transición Fuente* y este caso sólo se requiere que ocurra su evento asociado para poder ser disparada, de forma similar, cuando una transición no posee ningún lugar de salida se dice que es una *Transición Sumidero* y cuando se dispara sólo se remueven marcas de los lugares que entran a ella. En la Figura 7 se presenta una red donde la transición t1 es de tipo fuente y la t3 de tipo sumidero. De forma análoga, existen los *Lugares Fuentes* (que no están conectados a ninguna transición de entrada) y *Lugares Sumidero* (que no están conectados a ninguna transición de salida) [19].

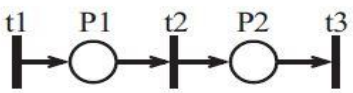


Figura 7. Transiciones fuente y sumidero *[19]*.

6.2. PROPIEDADES DE LAS RDP

- **RdP limitada**

Se dice que una red de Petri es k-limitada cuando el número de marcas en cada uno de los lugares de la red no excede un número finito k para cualquier marcado alcanzable a partir del marcado inicial. Además, una red está viva si es posible siempre disparar alguna transición de la red, sin importar que marcado se haya alcanzado ni cual sea la secuencia de disparo futura.

El máximo número de marcas para cada uno de los lugares de una red es su *Capacidad*, por lo que se dice que un lugar está limitado si su capacidad es finita.

- **RdP viva**

Una red de Petri se dice que es viva si, una vez alcanzado un marcado cualquiera desde M_0, siempre es posible disparar cualquier transición de la red mediante una secuencia progresiva y adecuada de disparos. Con frecuencia esta propiedad se relaciona directamente con la no existencia de puntos muertos que ocasionen la imposibilidad de disparar cualquier transición [19].

- **RdP reversible**

Una red de Petri es reversible o cíclica cuando después de una secuencia de marcado cualquiera, es posible volver al marcado inicial M_0.

- **RdP persistente**

Una red de Petri es persistente si dos transiciones cualesquiera que están sensibilizadas permanecen de igual forma hasta su respectivo disparo. Esto implica que, si se dispara una de las transiciones, entonces la otra continúa sensibilizada.

6.3. ARQUITECTURA DE COLAS PARA RDP

Las colas o filas se presentan de forma común en los sistemas de atención a usuarios. Como ejemplo de una arquitectura de colas, se muestra en la Figura 8 el modelo de un sistema de atención al cliente donde se posee dos tipos diferentes de servicios a prestar mediante el uso de tres ventanillas de atención [19].

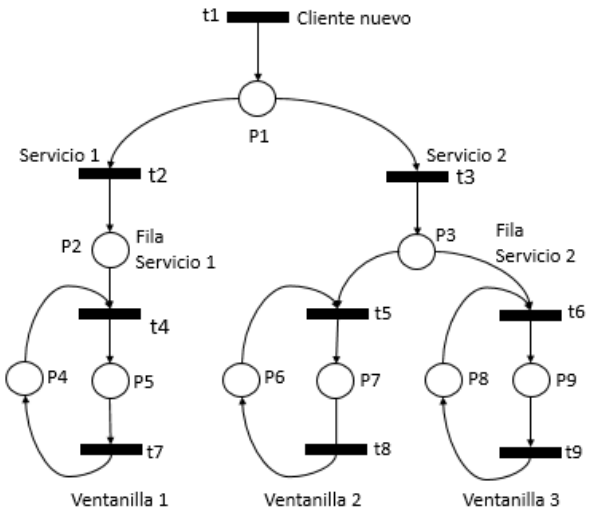


Figura 8. Arquitectura para colas *[19]*.

6.4. ANÁLISIS DE LAS REDES DE PETRI

6.4.1. Matrices de incidencia previa y posterior

Cada elemento que conforma la matriz de incidencia posterior, $\mathbb{C}^+$, para una red de Petri es el peso del arco que va desde la transición t_j al lugar de salida p_i y se define como $c_{ij}^+ = \beta(t_j, p_i)$. De lo anterior $\mathbb{C}^+ = \left[c_{ij}^+\right]$ con dimensiones $m \times n$, para una red con m lugares y n transiciones. En la matriz de incidencia previa, $\mathbb{C}^-$, cada uno de sus elementos representa el peso del arco que llega a la transición t_j proveniente desde el lugar de entrada p_i y se define como $c_{ij}^- = \alpha(p_i, t_j)$. De lo anterior $\mathbb{C}^- = \left[c_{ij}^-\right]$ con dimensiones $m \times n$ [19].

En general, la matriz de incidencia posterior representa el peso de los arcos de salida de cada una de las transiciones de la red, mientras que la matriz de incidencia previa representa el peso de los arcos de entrada a cada una de las transiciones. La figura siguiente se usará como ejemplo para el análisis.

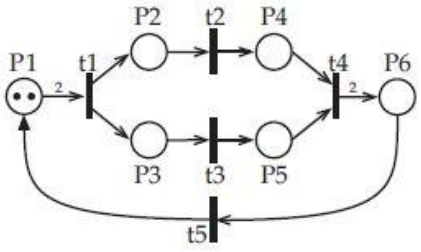


Figura 9. Red ejemplo *[19]*.

A continuación se muestra las matrices de incidencia previa ($\mathbb{C}^+$) e incidencia posterior ($\mathbb{C}^-$) para la red de la Figura 9.

$$\mathbb{C}^+ = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 2 & 0 \end{bmatrix} \qquad \mathbb{C}^- = \begin{bmatrix} 2 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

6.4.2. Matriz de incidencia

En las RdP la representación matricial se puede simplificar en una única matriz denominada *Matriz de incidencia* ($\mathbb{C}$), la cual se define como: $\mathbb{C} = \mathbb{C}^+ - \mathbb{C}^-$. Cada uno de los elementos que compone la matriz de incidencia, c_{ij}, es positivo para indicar la presencia de incidencia posterior, negativo para indicar la presencia de incidencia previa o cero para indicar la no conexión entre el lugar p_i y la transición t_j.

A continuación, se muestra la matriz de incidencia para la red de la Figura 9.

$$\mathbb{C} = \mathbb{C}^+ - \mathbb{C}^- = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 2 & 0 \end{bmatrix} - \begin{bmatrix} 2 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} -2 & 0 & 0 & 0 & 1 \\ 1 & -1 & 0 & 0 & 0 \\ 1 & 0 & -1 & 0 & 0 \\ 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 2 & -1 \end{bmatrix}$$

6.4.3. Ecuación de estado

En una red de Petri interesa conocer el marcado que se alcanza luego de realizar una cierta secuencia de disparo partiendo de un marcado inicial conocido. El vector de disparo, μ_k, es un vector con tantos elementos como transiciones tiene la red y con un valor de 1 en la posición de la transición disparada y cero en el resto de componentes [19].

En las redes de Petri con un marcado inicial conocido, existe una ecuación que determina el marcado alcanzado desde un marcado inicial dado un vector de disparo y se conoce como ecuación de estado, la cual está dada por:

$$M_k^T = M_{k-1}^T + \mathbb{C}_{\mu k} \tag{24}$$

Como esta ecuación entrega una forma de determinar un marcado posterior a partir de uno previo, entonces se puede decir sucesivamente que:

$$M_k^T = M_{k-2}^T + \mathbb{C}_{\mu k-1} + \mathbb{C}_{\mu k} = M_0^T + \mathbb{C}_{(\mu_1+\cdots+\mu_k)} \quad (25)$$

De la anterior ecuación se puede definir el vector de disparo

$$\sigma = \mu_1 + \cdots + \mu_k \quad (26)$$

Al ser reemplazada en la ecuación (*25*) se obtiene:

$$M_k^T = M_0^T + \mathbb{C}\sigma \quad (27)$$

Aplicando las ecuaciones anteriores a la red de la Figura 9, se puede determinar el marcado alcanzado luego de disparar las transiciones t1 y t2 así:

$$M_1^T = \begin{bmatrix}2\\0\\0\\0\\0\\0\end{bmatrix} + \begin{bmatrix}-2 & 0 & 0 & 0 & 1\\ 1 & -1 & 0 & 0 & 0\\ 1 & 0 & -1 & 0 & 0\\ 0 & 1 & 0 & -1 & 0\\ 0 & 0 & 1 & -1 & 0\\ 0 & 0 & 0 & 2 & -1\end{bmatrix}\begin{bmatrix}1\\1\\0\\0\\0\end{bmatrix} = \begin{bmatrix}2\\0\\0\\0\\0\\0\end{bmatrix} + \begin{bmatrix}-2\\0\\1\\1\\0\\0\end{bmatrix} = \begin{bmatrix}0\\0\\1\\1\\0\\0\end{bmatrix}$$

Si ahora se dispara las transiciones t3 y t4 el marcado alcanzado es:

$$M_1^T = \begin{bmatrix}0\\0\\1\\1\\0\\0\end{bmatrix} + \begin{bmatrix}-2 & 0 & 0 & 0 & 1\\ 1 & -1 & 0 & 0 & 0\\ 1 & 0 & -1 & 0 & 0\\ 0 & 1 & 0 & -1 & 0\\ 0 & 0 & 1 & -1 & 0\\ 0 & 0 & 0 & 2 & -1\end{bmatrix}\begin{bmatrix}0\\0\\1\\1\\0\end{bmatrix} = \begin{bmatrix}0\\0\\1\\1\\0\\0\end{bmatrix} + \begin{bmatrix}0\\0\\-1\\-1\\0\\2\end{bmatrix} = \begin{bmatrix}0\\0\\0\\0\\0\\2\end{bmatrix}$$

- **Determinación de la reversibilidad**

Una red de Petri es reversible si existe un vector anulador derecho, Γ, con todos sus elementos positivos para la matriz de incidencia de la red; esta definición establece que:

$$\mathbb{C}\Gamma = 0 \quad (28)$$

Siguiendo con el ejemplo:

$$\mathbb{C}\Gamma = \begin{bmatrix}-2 & 0 & 0 & 0 & 1\\ 1 & -1 & 0 & 0 & 0\\ 1 & 0 & -1 & 0 & 0\\ 0 & 1 & 0 & -1 & 0\\ 0 & 0 & 1 & -1 & 0\\ 0 & 0 & 0 & 2 & -1\end{bmatrix}\begin{bmatrix}\gamma_1\\ \gamma_2\\ \gamma_3\\ \gamma_4\\ \gamma_5\end{bmatrix} = 0 \qquad \begin{aligned} -2\gamma_1 + \gamma_5 &= 0\\ \gamma_1 - \gamma_2 &= 0\\ \gamma_1 - \gamma_3 &= 0\\ \gamma_2 - \gamma_4 &= 0\\ \gamma_3 - \gamma_4 &= 0\\ 2\gamma_4 - \gamma_5 &= 0\end{aligned}$$

De donde se obtiene $\gamma_1 = \gamma_2 = \gamma_3 = \gamma_4 = \pi$ y $\gamma_5 = 2\pi$. Con $\pi = 1$ el vector $\Gamma = [1,1,1,1,2]^T$, lo cual significa que se requiere exactamente disparar una vez las transiciones t1 a t4 y dos veces la transición t5 para partir y regresar al marcado inicial.

- **Determinación de la conservatividad**

Una red de Petri es conservativa si y solo si existe un vector anulador izquierdo, Δ, tal que:

$$\Delta^T \mathbb{C} = 0 \tag{29}$$

Determinando la conservatividad para el ejemplo:

$$\Delta\mathbb{C} = [\delta_1 \quad \delta_2 \quad \delta_3 \quad \delta_4 \quad \delta_5 \quad \delta_6] \begin{bmatrix} -2 & 0 & 0 & 0 & 1 \\ 1 & -1 & 0 & 0 & 0 \\ 1 & 0 & -1 & 0 & 0 \\ 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 2 & -1 \end{bmatrix} = 0$$

$$\begin{gathered} -2\delta_1 + \delta_2 + \delta_3 = 0 \\ -\delta_2 + \delta_4 = 0 \\ -\delta_3 + \delta_5 = 0 \\ -\delta_4 - \delta_5 + 2\delta_6 = 0 \\ \delta_1 - \delta_6 = 0 \end{gathered}$$

De donde se obtiene lo siguiente:

- $\delta_1 = \delta_6 = \pi_1$
- $\delta_3 = \delta_5 = \pi_2$
- $\delta_2 = \delta_4 = \pi_3$
- $2\pi_1 = \pi_2 + \pi_3$

Siendo $\pi_2 = \pi_3 = 1$, entonces el vector Δ= $[1,1,1,1,1,1]^T$. Los resultados anteriores hacen notar la realidad de la Conservatividad de la red, porque se interpretan como que el disparo de t2 retira el mismo número de marcas para los lugares P2 y P4, mientras que t3 hace lo propio para los lugares P3 y P5 y que t1 y t4 retiran la misma cantidad de marcas que agregan, como lo evidencia la ecuación $2\pi_1 = \pi_2 + \pi_3$

6.4.4. Redes de Petri Estocásticas

Las Redes de Petri Estocásticas (SPN) se obtienen asociando con cada transición de la red una variable aleatoria con una distribución probabilística que describa el retardo desde la habilitación hasta el disparo de dicha transición.

Como caso hipotético, se puede considerar una SPN y un marcado M en el cual múltiples transiciones están simultáneamente habilitadas. La transición que tiene asociado el retardo más breve disparará primero. La SPN alcanza un nuevo marcado M', donde algunas transiciones estuvieron habilitadas en el marcado anterior, pero que no fueron disparadas, por lo tanto, pueden estar habilitadas aún. De acuerdo con la propiedad de falta de memoria de las variables aleatorias, se obtiene una distribución de vida igual a la distribución del retardo de disparo. Lo que concluye que la actividad asociada con cada transición *recomienza* con cualquier nuevo marcado [20].

Una definición formal de una SPN es:

$$SPN = \{P, T, A, M_0, L\}$$

Donde P, T, A y M_0 se definen como antes y $L = (l_1, l_2, \dots, l_m)$ es el conjunto de tasas de retardos asociados con las transiciones.

PARTE III

RESULTADOS, ANÁLISIS Y CONCLUSIONES

CAPÍTULO 7.

ANÁLISIS DE LÍNEAS DE ESPERA A TRAVÉS DE RDP

Las RdP pueden mostrar lo que sucede con cada uno de sus elementos en un instante deseado, por ejemplo, ver cuántas veces se ha disparado una transición o cuántas marcas hay en un lugar. También permiten hacerlo a nivel global y saber en un instante cómo se encuentra la totalidad de la red, esto gracias a las matrices de incidencia y al método gráfico que permite hacer un análisis parte por parte de cada uno de los elementos de la red.

7.1. ANALOGÍA DE LÍNEAS DE ESPERA Y GRÁFICOS DE REDES DE PETRI

Para lograr hacer un correcto análisis de una RdP a una línea de espera, es necesario conocer su estructura y la función de cada componente para realizar una correspondiente analogía y entender sus limitaciones.

Cabe recordar que las transiciones en la RdP se dan para acciones instantáneas (toma de decisiones, entrar o salir de una acción a otra, etc.), los lugares se dan para ejecutar acciones que toman tiempo (mover algo, esperar, tiempo para tomar una decisión, etc.) y los arcos conectan acciones instantáneas con acciones que tardan tiempo en realizarse.

7.1.1. Componentes y parámetros de una línea de espera

Los componentes que son motivo de análisis en una línea de espera son:

- **Cliente**

El cliente es el actor principal de la línea de espera y pasa por todo el sistema de la cola, desde la entrada hasta la salida. Esta acción de moverse a través del sistema se ve reflejada en las marcas de la red. Dado esto, lo que para teoría de colas es *n*-clientes, para RdP será *n*-marcas.

- **Entrada al sistema**

La entrada al sistema es una acción transitoria que se produce cuando un cliente toma la decisión de ingresar al sistema. Para teoría de colas, la entrada al sistema o cola viene dada por el parámetro λ (tasa de llegadas), este parámetro es análogo a la tasa de disparos de la transición de entrada al sistema en RdP.

- **Cola**

La cola es un punto estacionario donde el cliente, que decidió entrar al sistema, debe esperar a que sea su turno para ingresar al servicio. Mencionado lo anterior, la cola se representa como un lugar en una RdP donde se acumulan *n*-marcas a espera que se dispare la transición de entrada al servicio.

- **Entrada al servicio**

Al igual que en la entrada al sistema, la entrada al servicio es una acción instantánea, por lo tanto, se representa como una transición en una RdP que se dispara si hay clientes en la cola y algún servidor está disponible.

- **Servicio**

El servicio es la parte final del sistema de las líneas de espera, ya que es donde el cliente recibe el servicio solicitado (comprar, pagar, solicitar, etc.) y es una acción estacionaria. Al ser una acción estacionaria, se representa como un lugar en una RdP con un límite de una marca y que se encuentra disponible si no hay marca en él.

- **Salida del sistema**

La salida del sistema también es una acción instantánea pero depende del tiempo que transcurre cuando un cliente entra al servicio hasta que lo finaliza, este parámetro se conoce como μ (tasa de servicio) en la teoría de colas, este parámetro es análogo a la tasa de disparo de la transición de salida en RdP.

7.1.2. Limitaciones físicas

Las limitaciones de las líneas de espera son: Límite de la cola, número de servidores y la disponibilidad del servicio. A continuación, se hará una analogía entre las limitaciones mencionadas con las Redes de Petri.

- **Límite de la cola**

El límite de la cola es una condición física o técnica (decidida por logística) que da información de cuál es el máximo de clientes que pueden estar en espera en un mismo instante y lugar. Las RdP pueden modelar restricciones con lugares en paralelo al lugar al que se le aplica la restricción, añadiendo una cantidad de marcas iguales al límite esperado.

- **Disponibilidad del servicio**

Si en el servicio hay un cliente, la cola no puede reducirse hasta que el cliente que lo ocupa finalice su servicio. Esta condición también se puede ver como un límite

para cada servidor y al igual que para el límite de la cola, se representa con un lugar en paralelo, pero solo con una marca.

- **Número de servidores**

El número de servidores es el comodín que se varía de acuerdo a la calidad del servicio que se quiere dar pero está limitada por aspectos económicos o físicos. En RdP, un servidor lo conforman: una transición de entrada, seguida de un lugar de servicio en paralelo con un lugar que indica la disponibilidad y finalizando con una transición de salida del sistema.

7.2. CONSTRUCCIÓN DE UNA RED DE PETRI PARA LÍNEAS DE ESPERA

De las definiciones y analogías en los numerales anteriores, a continuación, se realizan los siguientes gráficos de Petri equivalentes a líneas de espera.

7.2.1. Línea de espera con única cola y único servidor

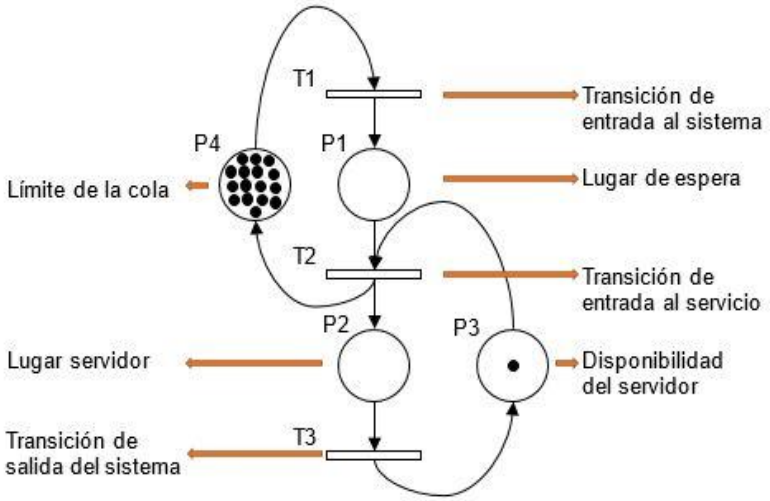


Figura 10. RdP para una línea de espera con única cola y único servidor. Fuente: Autor

La Figura 10 muestra que la forma de construir la RdP para un proceso de cola con única cola y único servidor consiste en asignar un lugar (P4) que representa el límite de la cola, un lugar (P1) como el lugar de espera o la cola como tal, un lugar (P2) como el lugar del servidor, es decir, donde el cliente recibe el servicio y un último lugar (P3) que representa la disponibilidad del servicio. En cuanto a las transiciones, éstas representan los tiempos de las diferentes etapas del servicio, así: T1 representa los tiempos de llegada de los clientes al sistema, T2 un tiempo de entrada al servicio y T3 el tiempo de salida del sistema. Todos los arcos tienen peso 1 porque en este sistema solo se atenderá un cliente a la vez.

- **Marcado inicial (M_0) y Matrices de incidencia ($\mathbb{C}^+$,$\mathbb{C}^-$,$\mathbb{C}$)**

  Se establece el marcado inicial con un límite de cola x

$$M_0 = [0 \quad 0 \quad 1 \quad x]$$

  Y las matrices de incidencia son:

$$\mathbb{C}^+ = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \qquad \mathbb{C}^- = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \qquad \mathbb{C} = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \\ 0 & -1 & 1 \\ 1 & 1 & 0 \end{bmatrix}$$

- **Propiedades**

A continuación se muestra la determinación de las propiedades para la red de la Figura 10.

- **Reversibilidad**

  Esta propiedad se evalúa con la intención de verificar si el sistema tiene la capacidad de regresar a su estado inicial. Usando la ecuación (28):

$$\mathbb{C}\Gamma = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \\ 0 & -1 & 1 \\ 1 & 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \gamma_3 \end{bmatrix} \qquad \begin{aligned} \gamma_1 - \gamma_2 &= 0 \\ \gamma_2 - \gamma_3 &= 0 \\ \gamma_3 - \gamma_2 &= 0 \\ \gamma_2 - \gamma_1 &= 0 \end{aligned}$$

  De donde se obtiene que: $\gamma_1 = \gamma_2 = \gamma_3 = \pi$ con $\pi = 1$. El vector anulador es $\Gamma = [1 \quad 1 \quad 1]$, lo cual significa que se requiere disparar exactamente de a una vez a las transiciones de T1 hasta T3 para partir y regresar al estado inicial. También se debe notar que cualquier valor de π es admisible para que se cumpla la reversibilidad. Para el caso de líneas de espera, quiere decir que si entran π clientes al sistema entonces π clientes deben de completar el circuito del sistema hasta salir del mismo.

- **Conservatividad**

  Usando la ecuación (29) se obtiene lo siguiente:

$$\Delta^T \mathbb{C} = [\delta_1 \quad \delta_2 \quad \delta_3 \quad \delta_4] \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \\ 0 & -1 & 1 \\ -1 & 1 & 0 \end{bmatrix}$$

$$\delta_1 - \delta_4 = 0$$
$$-\delta_1 + \delta_2 - \delta_3 + \delta_4 = 0$$
$$-\delta_2 + \delta_4 = 0$$

De donde se obtiene que: $\delta_1 = \delta_2 = \delta_3 = \delta_4 = \pi = 1$, el vector anulador es $\Delta = [1 \quad 1 \quad 1 \quad 1]$. Esto significa que las tres transiciones cuando se disparan absorben y entregan la misma cantidad de marcas, por lo tanto, la red es conservativa.

- **Limitada**

  Por definición se sabe que si la red es conservativa también es limitada. Si el análisis se hace para una cola de capacidad ilimitada entonces x es un valor lo suficientemente grande como para que la cola no la alcance en ningún instante de tiempo. Para este caso se dice que la RdP es x limitada.

- **Viva**

  La red es viva porque independiente del marcado que esta posea, siempre es posible disparar una transición.

➢ **Ejemplo numérico**

A continuación se analiza un sistema con única cola y único servidor de disciplina FIFO, sin cola al inicio, con una tasa de llegadas (λ) de 1 cliente cada 4 min y una tasa de servicio (μ) de 1 cliente cada 6 min, sin límite de cola. El análisis se realiza para un período de 1 hora.

El sistema para la Red de Petri es como el de la Figura 10 y se sabe por el análisis hecho anteriormente que la red es reversible (de acuerdo a una secuencia de disparos puede regresar a su estado inicial), conservativa (la totalidad de marcas en el sistema se conserva para cualquier disparo posible) y limitada (la cantidad máxima de marcas en un lugar es un número definido y no variable).

▪ **Factor de utilización**

El factor de utilización indica que tan cargado está el sistema con respecto a su carga máxima de trabajo por unidad de tiempo.

$$\rho = \frac{1/\lambda}{1/\mu} = \frac{0.25}{0.1667} = 1.5$$

Esto quiere decir que el sistema está 50 puntos porcentuales por encima de su límite máximo.

- **Vector de disparo**

El vector de disparo brinda información acerca de cuántas veces se han disparado todas las transiciones en un instante de análisis.

- **Disparo transición T1**

  La transición T1 es la transición de entrada al sistema y se dispara de acuerdo a la tasa de llegadas, dado que la entrada de cada cliente significa el disparo de la misma.

$$\text{Disparos llegada } T1\ (0-t) = [\lambda_p * t]$$
$$\text{Disparos llegada } T1\ (0-60) = [0.25 * 60]$$
$$\text{Disparos llegada } T1\ (0-60) = 15$$
$$\text{Disparos llegada } T1 = 15$$

  El resultado anterior significa que la transición T1 se ha disparado 15 veces, o bien, que al sistema han entrado 15 clientes a los 60 minutos.

- **Disparo transición T3**

  La transición T3 es la transición de salida del sistema y se dispara de acuerdo a la tasa de servicio, dado que la salida de cada cliente significa el disparo de la transición T3.

$$\text{Disparos salida T3}\ (0-\text{t}) = \left[\mu_\text{p} * \text{t} - \frac{\mu_\text{p}}{\lambda_\text{p}}\right]$$
$$\text{Disparos salida T3}\ (0-60) = \left[0.1667 * 60 - \frac{0.1667}{0.25}\right]$$
$$\text{Disparos salida T3}\ (0-60) = 9.333$$
$$\text{Disparos salida T3}\ (0-60) = 9$$

  Esto significa que la transición T3 se ha disparado 9 veces, o bien, que han salido 9 clientes a los 60 minutos.

- **Disparo transición T2**

  La transición T2 es la transición de entrada al servicio y para su cálculo se analiza que, si han salido x clientes y en el servicio siempre hay un clientes, a la ecuación de salida del sistema solo basta con sumarle 1 para obtener los disparos de la transición T2.

$$\text{Disparos salida T2}\ (0-\text{t}) = \left[\mu_\text{p} * \text{t} - \frac{\mu_\text{p}}{\lambda_\text{p}}\right] + 1$$
$$\text{Disparos salida T2}\ (0-\text{t}) = [9.333] + 1$$

$$\text{Disparos salida T2 } (0 - \text{t}) = 10$$

Lo que significa que la transición T2 se ha disparado 10 veces, o bien, que han entrado 10 clientes al servicio una vez transcurrido 60 minutos.

Dados los resultados anteriores, el vector de disparo queda de la siguiente manera:

$$\sigma = \begin{bmatrix} 15 \\ 9 \\ 10 \end{bmatrix}$$

- **Marcado inicial [M_0]**

El marcado inicial son las marcas que contiene cada lugar antes de que el sistema inicie. Para el ejercicio actual se trata de no dar un límite a la cola, para ello se asume 100 marcas en el lugar que representa el límite de la cola, tal que nunca se pueda alcanzar dicho número de clientes. Dado lo anterior, el marcado inicial quedaría:

$$M_0 = [0 \quad 0 \quad 1 \quad 100]$$

Se aplica la ecuación (27) para obtener el marcado o estado del sistema pasados 60 minutos.

$$M_1^T = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 100 \end{bmatrix} + \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \\ 0 & -1 & 1 \\ 1 & 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} 15 \\ 10 \\ 9 \end{bmatrix}$$

$$\boldsymbol{M_1} = [5 \quad 1 \quad 0 \quad 95]$$

El resultado obtenido quiere decir que se encuentran 5 marcas en el lugar de espera en la cola lo que se puede ver como 5 clientes en espera de ser atendidos, 1 marca en el lugar de servicio significa que hay un cliente recibiendo el servicio, 0 marcas en el lugar que representa el límite del servidor, esto significa que el servidor se encuentra ocupado y 95 marcas en el lugar correspondiente al límite de la cola que se puede interpretar como 95 clientes potenciales. Todo esto se puede apreciar en la Figura 11, la cual es una representación en RdP del estado del sistema pasados 60 minutos de análisis.

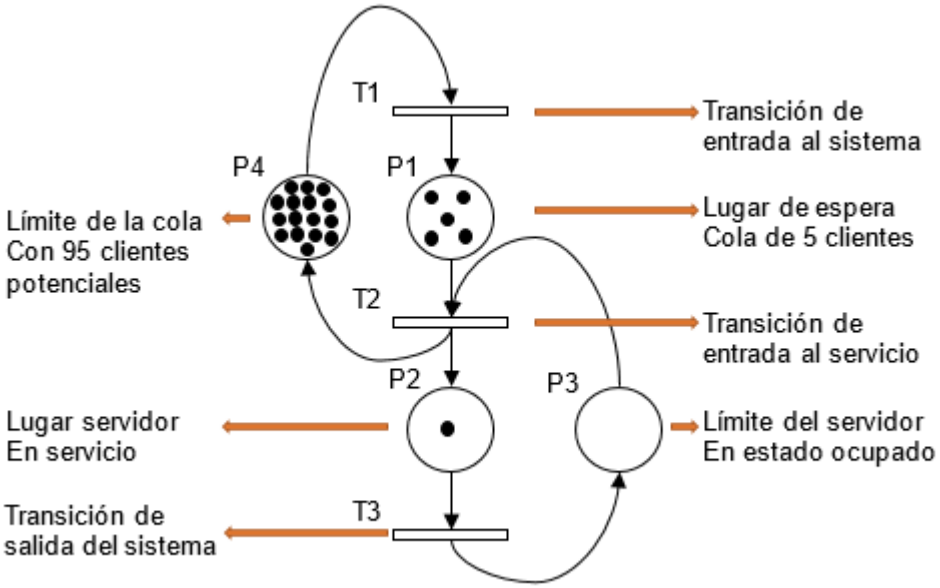


Figura 11. Estado de la RdP pasados 60 minutos de análisis. Fuente: Autor

Para el cálculo de las medidas de eficiencia se utilizan las ecuaciones (4) para el cálculo del tamaño del sistema y tamaño de la cola y (5) para hallar el tiempo del n-ésimo cliente en cola, además utilizando la relación $W = W_q + \frac{1}{\mu}$ se encuentra el tiempo del n-ésimo cliente en el sistema.

- **Clientes en el sistema**

$$L(t) = [\lambda * t] - \left[\mu * t - \frac{\mu}{\lambda}\right]$$

$$L(t) = [0.25 * 60] - \left[0.1667 * 60 - \frac{0.1667}{0.25}\right]$$

$$L(t) = 15 - 9 = 6 \text{ clientes}$$

- **Tamaño de la cola**

$$L_q(t) = [\lambda * t] - \left[\mu * t - \frac{\mu}{\lambda}\right]$$

$$L_q(t) = 15 - 9 - 1$$

$$L_q(t) = 5 \text{ clientes}$$

- **Tiempo del n-ésimo cliente en cola**

$$W_q = \left(\frac{1}{\mu} - \frac{1}{\lambda}\right)(T-1)$$

$$W_q = \left(\frac{1}{0.1667} - \frac{1}{0.25}\right)(15-1)$$

$$W_q = 28 \text{ minutos}$$

- **Tiempo del n-ésimo cliente en el sistema**

$$W = W_q + \frac{1}{\mu}$$

$$W = 28 \text{ min} + 6 \text{ min}$$

$$W = 34 \text{ minutos}$$

- **Gráfica de comportamiento del sistema en el tiempo**

Si se hace una suposición de que en el minuto 60 el sistema cierra la entrada, lo que quiere decir que $\lambda = 0$, el sistema tendría el comportamiento que se muestra en la Figura 12 en donde la línea de color azul representa el número de clientes que ingresan al sistema hasta los 60 minutos, mientras que la línea de color naranja hace referencia a los clientes que salen.

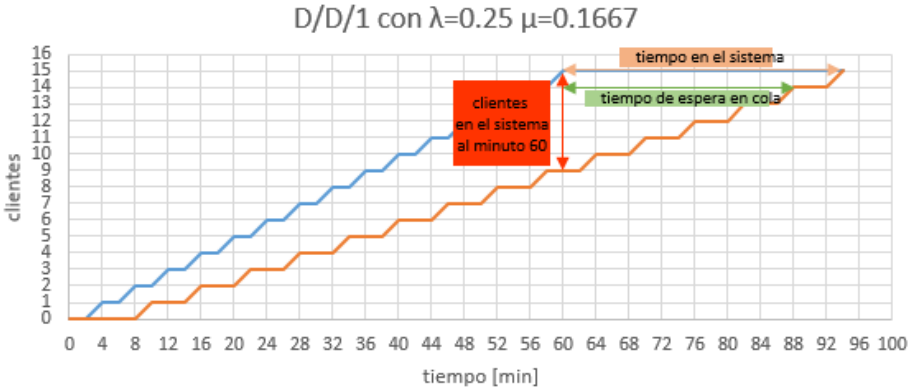


Figura 12. Gráfica del comportamiento del sistema D/D/1

De la gráfica anterior se puede observar que las distancias brindan información del tamaño del sistema y los tiempos de espera. La diferencia vertical entre las líneas

azul y naranja entrega información acerca de la cantidad de clientes que hay en el sistema, dado lo anterior, se puede decir que la distancia máxima vertical hace referencia a la máxima cantidad de clientes que hay en el sistema tras el tiempo de análisis. Las diferencias horizontales muestran el tiempo que tarda un cliente en la cola, de igual forma, el máximo valor de las diferencias horizontales representa el máximo tiempo de un cliente en el sistema. También se puede extraer de este gráfico que tras cerrar el sistema a los 60 minutos de estudio, se requiere de otros 32 minutos para poder terminar de servir a todos los clientes de la cola.

El modelo usado en este numeral es el sistema más básico para simular un sistema de colas. Si bien funciona correctamente para un sistema en donde tanto los tiempos de llegadas y de servicio se conocen y son constantes, los sistemas de atención en donde los clientes son personas no tienen ese comportamiento, ya que los tiempos se comportan de forma aleatoria. Lo anterior hace a este sistema ineficiente de representar, por ejemplo una cola en un banco o en un supermercado. La otra limitación que presenta este sistema es su capacidad de servicio, porque al tener un solo servidor, no tiene la capacidad suficiente para servir a los clientes ante un crecimiento indefinido de la cola.

El modelo presentado a continuación, cuenta con la con las herramientas para resolver la falencia de este modelo en cuanto a la cantidad de servidores.

7.2.2. Línea de espera con única cola y n servidores

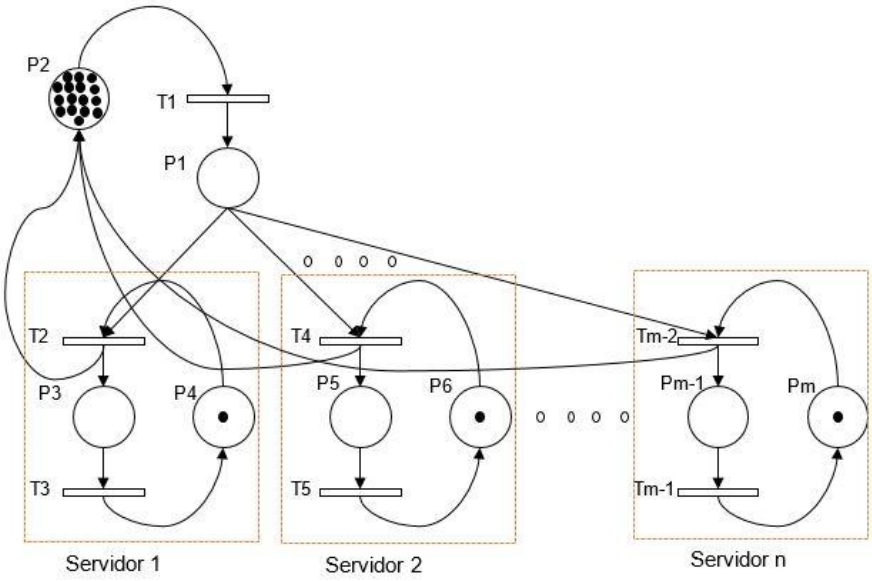


Figura 13. RdP para una línea de espera con única y n servidores. Fuente: Autor

La Figura 13 muestra un sistema de única cola con varios servidores. Este sistema resuelve los problemas mencionados previamente en cuanto a capacidad de

servicio, porque al haber varios servidores el sistema se puede descongestionar más rápido y evita que la cola crezca de forma descontrolada. Cuenta con un lugar (P2) que simula la población que puede entrar al sistema. Una transición (T1) como la transición de entrada al sistema, un lugar (P1) que representa la fila única. Como son varios servidores en paralelo, están construidos de manera idéntica, una transición de entrada al servicio, después de esto un lugar que representa la prestación del servicio y por último una transición que simula la salida del servicio. Paralelo a esto, hay un lugar que representa la disponibilidad del servidor.

- **Marcado inicial (M_0) y Matrices de incidencia ($\mathbb{C}^+$,$\mathbb{C}^-$,$\mathbb{C}$)**

  Se establece el marcado inicial con un límite de cola x

$$M_0 = [0 \quad x \quad 0 \quad 1 \quad 0 \quad 1 \quad \cdots \quad 0 \quad 1]$$

  Las matrices de incidencia son las siguientes:

$$\mathbb{C}^+ = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & \cdots & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 1 \end{bmatrix} \quad \mathbb{C}^- = \begin{bmatrix} 0 & 1 & 0 & 1 & 0 & \cdots & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & \cdots & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & \cdots & 1 & 0 \end{bmatrix}$$

$$\mathbb{C} = \begin{bmatrix} 1 & -1 & 0 & -1 & 0 & \cdots & -1 & 0 \\ -1 & 1 & 0 & 1 & 0 & \cdots & 1 & 0 \\ 0 & 1 & -1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & \cdots & 0 & 0 \\ 0 & 0 & 0 & -1 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \cdots & 1 & -1 \\ 0 & 0 & 0 & 0 & 0 & \cdots & -1 & 1 \end{bmatrix}$$

➢ **Propiedades**

- **Reversibilidad**

  Esta propiedad se evalúa con la intención de verificar si el sistema tiene la capacidad de regresar al estado inicial. Usando la ecuación (28):

$$\mathbb{C}\Gamma = \begin{bmatrix} 1 & -1 & 0 & -1 & 0 & \cdots & -1 & 0 \\ -1 & 1 & 0 & 1 & 0 & \cdots & 1 & 0 \\ 0 & 1 & -1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & \cdots & 0 & 0 \\ 0 & 0 & 0 & -1 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \cdots & 1 & -1 \\ 0 & 0 & 0 & 0 & 0 & \cdots & -1 & 1 \end{bmatrix} \cdot \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \gamma_3 \\ \gamma_4 \\ \gamma_5 \\ \vdots \\ \gamma_{m-2} \\ \gamma_{m-1} \end{bmatrix}$$

$$\begin{gathered} \gamma_1 - \gamma_2 - \gamma_3 - \cdots - \gamma_{m-2} = 0 \\ -\gamma_1 + \gamma_2 + \gamma_3 + \cdots + \gamma_{m-2} = 0 \\ \gamma_2 - \gamma_3 = 0 \\ -\gamma_2 - \gamma_3 = 0 \\ \gamma_4 - \gamma_5 = 0 \\ -\gamma_4 + \gamma_5 = 0 \\ \vdots \\ \gamma_{m-1} - \gamma_{m-2} = 0 \\ \gamma_{m-2} - \gamma_{m-1} = 0 \end{gathered}$$

De donde se obtiene que: $\gamma_1 = \gamma_2 = \gamma_3 = \gamma_4 = \gamma_5 = \cdots = \gamma_{m-2} = \gamma_{m-1} = \pi$, y con $\pi = 1$, entonces el vector anulador es $\Gamma = [1\ 1\ 1\ 1\ 1 \ldots 11]$. Esto significa que se requiere disparar exactamente de a una vez a las transiciones desde T1 hasta Tm-1 para partir y regresar al estado inicial; se debe notar que cualquier valor de π es admisible para que se cumpla la reversibilidad. En el caso de líneas de espera, lo anterior quiere decir que si entran π clientes, entonces π clientes deben de completar el circuito del sistema hasta salir del mismo, sin importar el servidor que utilicen.

- **Conservatividad**

  Usando la ecuación (29):

$$\Delta\mathbb{C} = [\delta_1 \quad \delta_2 \quad \delta_3 \quad \delta_4 \quad \delta_5 \quad \cdots \quad \delta_{m-1} \quad \delta_m] \begin{bmatrix} 1 & -1 & 0 & -1 & 0 & \cdots & -1 & 0 \\ -1 & 1 & 0 & 1 & 0 & \cdots & 1 & 0 \\ 0 & 1 & -1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & \cdots & 0 & 0 \\ 0 & 0 & 0 & -1 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \cdots & 1 & -1 \\ 0 & 0 & 0 & 0 & 0 & \cdots & -1 & 1 \end{bmatrix}$$

$$\delta_1 - \delta_2 - \delta_4 - \delta_{m-1} = 0$$
$$-\delta_1 + \delta_2 + \delta_4 + \delta_{m-1} = 0$$
$$\delta_2 - \delta_3 = 0$$
$$-\delta_2 + \delta_3 = 0$$
$$\delta_4 - \delta_5 = 0$$
$$-\delta_4 + \delta_5 = 0$$
$$\vdots$$
$$\delta_{m-1} - \delta_m = 0$$
$$-\delta_{m-1} + \delta_m = 0$$

De lo anterior se obtiene: $\delta_1 = \delta_2 = \delta_3 = \delta_4 = \delta_5 = \delta_{m-1} = \delta_m = \pi = 1$, y el vector anulador es $\Delta = [1\ 1\ 1\ 1\ 1\ \ldots 1\ 1]$. Lo anterior significa que todas las transiciones, cuando se disparan, absorben la misma cantidad de marcas que las que entregan al sistema, por lo tanto, se conserva el marcado.

- **Limitada**

  Se sabe que si la red es conservativa también es limitada. Si el análisis se hace para una cola de capacidad ilimitada, entonces x es un valor lo suficientemente grande como para que la cola no lo alcance en ningún momento. Para el caso de ejemplo, la red es x limitada.

➢ **Ejemplo numérico**

Basado en el ejemplo anterior, a partir del minuto 60, se mantiene la tasa de llegadas y se añade un servidor más sistema con el fin de disminuir la cola más rápido, con lo cual se obtienen los siguientes resultados:

▪ **Factor de utilización**

$$\rho = \frac{1/\lambda}{c * 1/\mu} = \frac{0.25}{2 * 0.1667}$$
$$\rho = \frac{0.25}{0.3334}$$
$$\rho = 0.75$$

El anterior resultado indica que el sistema está cargado al 75%, para este sistema determinístico esto significa que la cola se disminuirá hasta no existir. Dado lo anterior, se calcula en cuánto tiempo disminuirá la cola. Despejando t de la ecuación (4) se obtiene:

$$L(t) = L_i(60) + [\lambda * t] - \left[\mu * t - \frac{\mu}{\lambda}\right] - c$$

$$t = \frac{L(t) - L_i(60) + c - \frac{\mu}{\lambda}}{(\lambda - \mu)}$$

$$t = \frac{0 - 6 + 2 - \frac{0.3334}{0.25}}{(0.25 - 0.3334)} = 64 \text{ min}$$

El resultado anterior quiere decir que al minuto 124 (60+64), la cola se reduce a 0 clientes y dado que el factor de utilización es menor a 1, se puede asegurar que no hay fila y que cada cliente entra directamente al servicio.

- **Vector de disparo**

  - **Disparo transición T1**

    Para este caso se calcula de igual manera que el anterior ejercicio, teniendo en la cuenta que el análisis se hace hasta el minuto 124 con tasa de llegadas igual para todo t.

$$\text{Disparos llegada T1 } (0 - \text{t}) = [\lambda p * t]$$
$$\text{Disparos llegada T1 } (0 - 124) = [0.25 * 124]$$
$$\text{Disparos llegada T1 } (0 - 124) = 31$$
$$\text{Disparos llegada T1 } = 31$$

    Lo que significa que la transición T1 se dispara 31 veces, o bien, que al sistema entran 31 clientes en total pasados los 124 minutos.

  - **Disparo transiciones T3 y T5**

    En este caso existen dos transiciones de salida del sistema y cada una se dispara con la misma tasa de servicio, por esto, se calcula de la misma manera que el ejemplo anterior, con la diferencia de que se multiplica la tasa de servicio por dos y además se calcula a partir del minuto 60, ya que la tasa de servicio del sistema cambia. Todo lo que se llevaba hasta ese momento se suma a la transición T3.

$$\text{Disparos Salida T3 y T5 } (60 - \text{t}) = [2 * \mu * t - \frac{\mu}{\lambda}]$$
$$\text{Disparos Salida T3 y T5 } (60 - 124) = [0.3334 * 64 - \frac{0.3334}{0.25}]$$
$$\text{Disparos Salida T3 y T5 } (60 - 124) = [20\ ]$$
$$\text{Disparos Salida T3 y T5 } (60 - 124) = 20$$

    Se puede ver a través del resultado anterior que la transición T3 y T5 entre el minuto 60 y el minuto 124 se disparan 20 veces, o dicho de otra forma, salen 20 clientes del sistema entre el minuto 60 y 124. Los disparos hechos desde

0 hasta 124 minutos por cada una de las transiciones quedan de la siguiente manera:

$$\mathrm{T3} = 19 \;\; ; \;\; \mathrm{T5} = 10$$

- **Disparo transiciones T2 y T4**

De igual manera que el ejercicio anterior, estas transiciones se deben haber disparado una vez más ya que son instantáneas y se refieren a la entrada al servidor.

$$\mathrm{Disparos\ Salida\ T2}\ (0-124) = 20$$
$$\mathrm{Disparos\ Salida\ T4}\ (0-124) = 11$$

Dado los resultados anteriores, el vector de disparo queda:

$$\sigma = \begin{bmatrix} 31 \\ 20 \\ 19 \\ 11 \\ 10 \end{bmatrix}$$

- **Marcado inicial** $[\mathbf{M_o}]$

Este marcado viene dado por como terminó el sistema a los 60 min más la adición del nuevo servidor. Para el ejercicio actual se trata de no dar un límite a la cola, para ello se asume 100 marcas en el lugar del límite de la cola tal que nunca se pueda alcanzar este número de clientes en la cola. Dicho lo anterior el marcado inicial es:

$$M_o = [0 \quad 100 \quad 0 \quad 1 \quad 0 \quad 1]$$

- **Marcado pasados 60 minutos** $[\mathbf{M_o}]$

Aplicando la ecuación (27) y trayendo la matriz de incidencia C que pertenece a este sistema, se encuentra el marcado o estado del sistema para el instante en análisis:

$${M_2}^T = \begin{bmatrix} 0 \\ 100 \\ 0 \\ 1 \\ 0 \\ 1 \end{bmatrix} + \begin{bmatrix} 1 & -1 & 0 & -1 & 0 \\ -1 & 1 & 0 & 1 & 0 \\ 0 & 1 & -1 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & -1 & 1 \end{bmatrix} \cdot \begin{bmatrix} 31 \\ 20 \\ 19 \\ 11 \\ 10 \end{bmatrix}$$

$$M_2 = [0 \quad 100 \quad 1 \quad 0 \quad 1 \quad 0]$$

- **Red de Petri pasados los 124 minutos de análisis**

En la Figura 14 se puede observar que pasados 124 minutos, el sistema se encuentra sin clientes en cola y con un cliente en cada servidor, esto debido a que se agregó un segundo servidor en el minuto 60. Y así, si se adiciona más servidores se puede decir con certeza que la cola disminuirá más rápido. A partir de esto se puede encontrar soluciones prácticas para un sistema determinístico en el cual se desea que la cola no supere un determinado límite y se usen los recursos de una manera óptima.

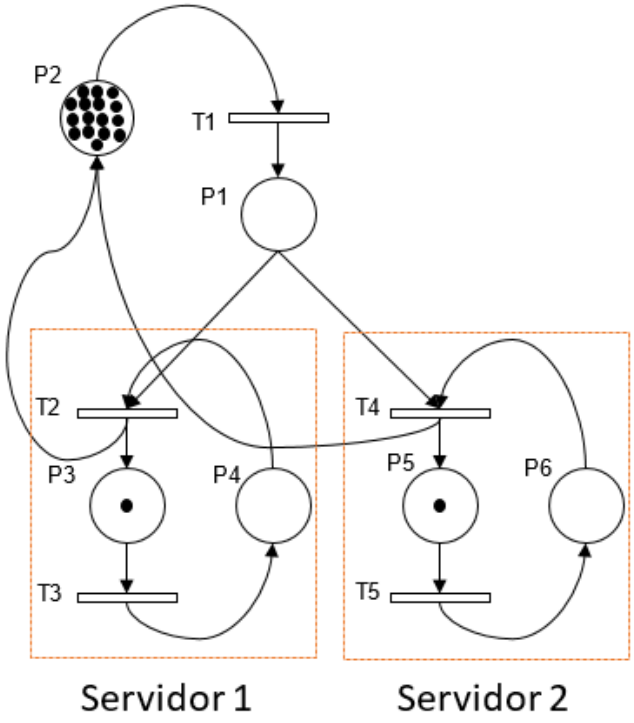


Figura 14. Estado de la RdP pasado 124 minutos

La Figura 15 relaciona la entrada de clientes (línea azul) con la salida de los mismos (línea naranja), de allí se pueden obtener las medidas de eficiencia como la longitud de la cola y tamaño del sistema que se calculan con la diferencia vertical entre la línea azul y la línea naranja, de igual forma se puede obtener el tiempo en la cola y tiempo total en el sistema calculando la diferencia horizontal de la línea azul y la línea naranja.

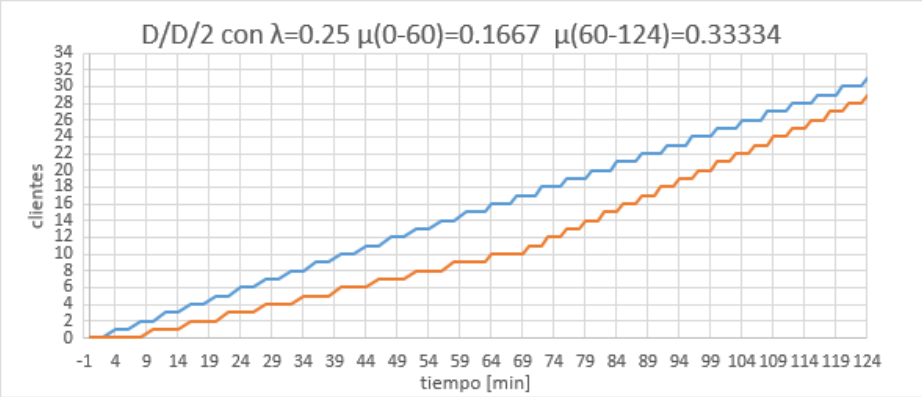


Figura 15. Gráfica de comportamiento del sistema en el tiempo del sistema D/D/2.

Este tipo de análisis se puede realizar para líneas de producción en donde los tiempos son deterministas y se tienen ciertas características que obligan al sistema a no superar una longitud de cola o tiempo de espera.

Este sistema, igual que el anterior (D/D/1) comparte la dificultad de poder simular sistemas de atención al público en donde los clientes sean personas por el comportamiento de llegada de los mismos que ya fue mencionado antes, por ello, se busca un sistema en donde los tiempos de llegada y de servicio sean aleatorios.

- **Ejercicio sistema M/M/1**

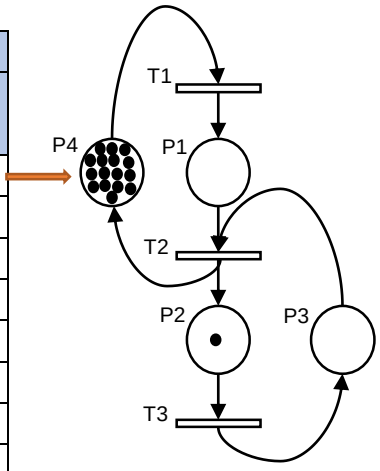
Este es un sistema con tasa de llegadas y tiempo de servicio probabilísticas. Como ya se observó en los anteriores ejercicios, el objetivo es hallar la forma de describir los disparos de cada transición. En este caso los disparos de las transiciones de entrada al sistema y salida del mismo están dadas por funciones de probabilidad, por lo cual es necesario la simulación de este sistema en las RdP para poder realizar el respectivo análisis.

A continuación se simula un ejercicio para el modelo M/M/1 por medio RdP con λ promedio de 0.33 minutos y μ promedio de 0.37. Los datos tanto para los tiempos entre llegadas como para los tiempos de servicio siguen una distribución exponencial. Las transiciones T1 y T3 tienen una frecuencia de disparo exponencial mientras que T2 se trabajó instantánea.

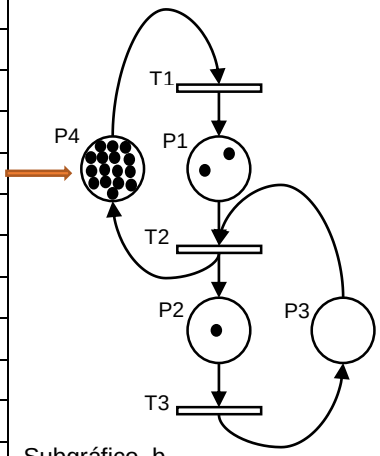
- **Simulación en RdP**

A continuación se muestra la Figura 16 que corresponde a la descripción de la simulación por Redes de Petri y se compone por una tabla en la parte izquierda y 4 subgráficos en la parte derecha. La tabla se distribuye en cuatro columnas, una llamada *Minuto* que evidencia el tiempo transcurrido y otras tres columnas que muestra los disparos por cada una de las transiciones a medida que avanza el tiempo. En la derecha se aprecia cuatro subgráficos (a, b, c, d), que en conjunto con la tabla describen el estado del sistema.

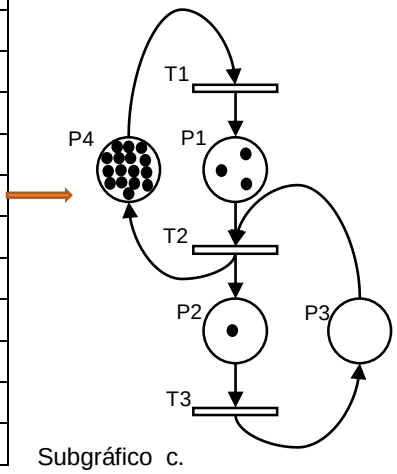
	DISPAROS		
Minuto	**T1 Exp**	**T2 Inst**	**T3 Exp**
2.31	1	1	
4.27	2		
6.35	3		
6.48		2	1
8.28	4		
10.91		3	2
10.92	5		
13.36		4	3
14.63	6		
18.49	7		
18.97		5	4
21.07		6	5
21.95	8		
24.80		7	6
24.91	9		
26.84		8	7
29.33		9	8
29.44	10		
31.73	11		
32.50		10	9
33.68	12		
34.34		11	10
35.50	13		
39.06	14		
40.07		12	11
42.92	15		
43.40		13	12
45.67	16		
46.01		14	13
47.42		15	14
48.43	17		
53.34		16	15
53.96	18		
54.25		17	16



Subgráfico a.

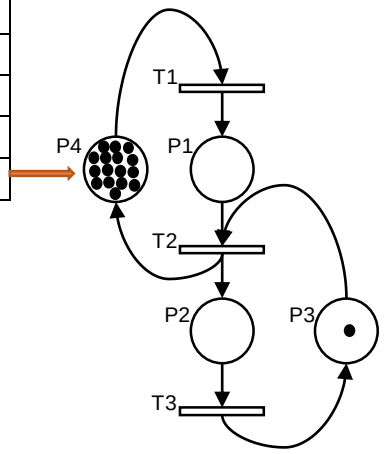


Subgráfico b.



Subgráfico c.

55.68		18	17
56.90			18
56.98	19	19	
59.04			19
60.23	20	20	
61.99			20
64.45	21	21	
67.56			
68.22			21



Subgráfico d.

Figura 16. Gráficos simulación RdP ejemplo M/M/1.

Como se observa al final de la tabla todas las transiciones se han disparado 21 veces, lo que corresponde a que 21 clientes entraron al sistema y que 21 clientes fueron atendidos. Los subgráficos de la dejan ver cómo se encuentra el sistema en ese instante de tiempo. El *Subgráfico a,* representa la entrada del primer cliente al sistema, por esto sigue directo al servicio y por esta razón se encuentra una marca en el lugar P2 (lugar de servicio); la situación del Subgráfico b deja ver que se acumulan dos clientes en la fila y se encuentra una persona siendo atendida; el Subgráfico c muestra que se encuentra tres personas en la cola y hay una persona siendo atendida y por último el Subgráfico d muestra que ya todos los clientes se atendieron y por lo tanto el servidor se encuentra desocupado y disponible.

- **Resultados de la simulación en RdP**

A través de la Red de Petri se puede obtener los resultados de las medidas de eficiencia como: W_q,W, L_q,L. La Tabla 1 muestra los tiempos de duración en la cola (W_q), es decir, el tiempo que se demoró desde que entró al sistema hasta que entró al servicio y tiempos de duración en el sistema (W) el cual es el tiempo total. Todo esto está debidamente identificado para cada uno de los clientes.

Tabla 1. Tiempos de duración en cola y en el sistema.

Cliente	Wq [min]	W [min]	Cliente	Wq [min]	W [min]	Cliente	Wq [min]	W [min]
1	0.00	4.17	8	4.89	7.38	15	4.5	10.43
2	2.21	6.64	9	4.42	7.59	16	7.68	8.59
3	4.56	7.01	10	3.06	4.90	17	5.83	7.25
4	5.08	10.69	11	2.61	8.34	18	1.72	2.94
5	8.04	10.15	12	6.39	9.72	19	0.00	2.06
6	6.45	10.18	13	7.89	10.51	20	0.00	1.76
7	6.31	8.35	14	6.95	8.35	21	0.00	3.77

La Tabla 2 muestra la cantidad de clientes que se encuentran en la cola (L_q) y los que hay en total en el sistema (L), cada vez que se dispara una transición.

Tabla 2. Tamaño del sistema y tamaño de la cola.

Minuto	Total en sistema	Total en cola	Minuto	Total en sistema	Total en cola	Minuto	Total en sistema	Total en cola
2.31	1	0	24.91	3	2	46.01	3	2
4.27	2	1	26.84	2	1	47.42	2	1
6.35	3	2	29.33	1	0	48.43	3	2
6.48	2	1	29.44	2	1	53.34	2	1
8.28	3	2	31.73	3	2	53.96	3	2
10.91	2	1	32.5	2	1	54.25	2	1
10.92	3	2	33.68	3	2	55.68	1	0
13.36	2	1	34.34	2	1	56.9	0	0
14.63	3	2	35.5	3	2	56.98	1	0
18.49	4	3	39.06	4	3	59.04	0	0
18.97	3	2	40.07	3	2	60.23	1	0
21.07	2	1	42.92	4	3	61.99	0	0
21.95	3	2	43.4	3	2	64.45	1	0
24.8	2	1	45.67	4	3	68.22	0	0

Cabe notar que a partir del minuto 55.68 los valores para los clientes en la cola es cero, lo que significa que la cola desaparece y que los clientes llegan directamente a los servidores para ser atendidos.

A continuación, en la Tabla 3 se observa un resumen general de los resultados obtenidos con la simulación en Redes de Petri.

Tabla 3. Resumen general del sistema.

Max W [minutos]	Max Wq [minutos]	Max L [clientes]	Max Lq [clientes]	ρ	λ [minutos]	μ [minutos]
11	8	4	3	90%	0.33	0.37

En la Tabla 4 se muestra la frecuencia con la que se repite un dato dentro de lo obtenido para las medidas de eficiencia.

Tabla 4. Frecuencias para las medidas de eficiencia.

Lq		**L**		**W**		**Wq**	
n	**frecuencia de n**	**n**	**frecuencia de n**	**t [min]**	**Frecuencia de t**	**t [min]**	**Frecuencia de t**
0	24%	0	10%	0.00	0%	0	14%
1	31%	1	14%	2.14	10%	1.61	5%
2	36%	2	31%	4.28	14%	3.22	19%
3	10%	3	36%	6.41	5%	4.83	14%
		4	10%	8.55	38%	6.43	24%
				10.69	33%	8.04	24%

La Tabla 4 deja ver una concordancia con la Tabla 3, ya que justamente los datos con mayor frecuencia son los datos máximos que se obtuvieron de las medidas de eficiencia.

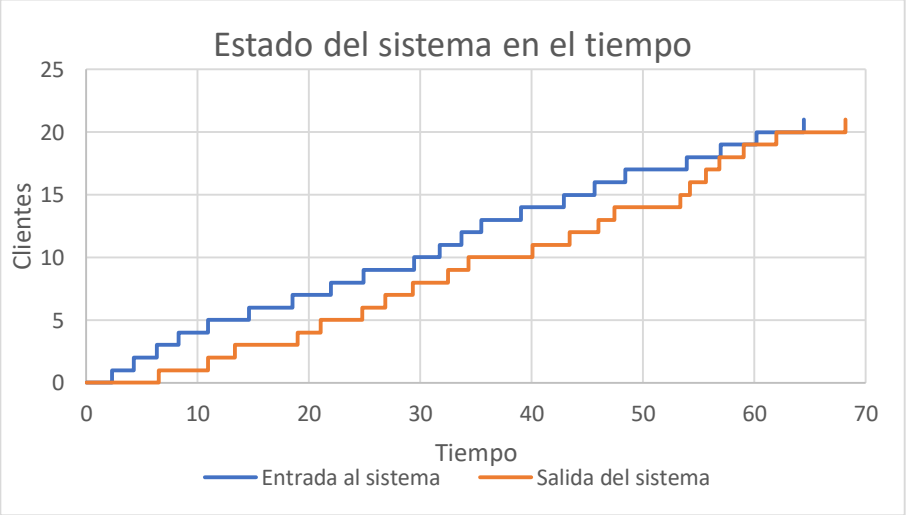


Figura 17. Gráfica de entradas y salidas del sistema en el tiempo de análisis.

La Figura 17 muestra el comportamiento del sistema en el tiempo, mediante la entrada y salida de los clientes. Se puede notar que en ciertos momentos las líneas

azul y naranja se unen, esto quiere decir que el número de entradas es igual al número clientes que salen del sistema haciendo que la cola deje de existir.

- **Resultados obtenidos a través de la Teoría de Colas**

Implementando las fórmulas (20), (21), (22) y (23) de teoría de colas, los resultados son los mostrados a continuación. Es pertinente agregar que los resultados obtenidos fueron resultados máximos, los cuales son comparables con los resultados de la Tabla 4.

Tabla 5. Resultados obtenidos por Teoría de Colas

W [minutos]	Wq [minutos]	L [clientes]	Lq [clientes]	Factor de utilización (ρ)
25	22.3	8	7	90%

Los datos obtenidos en la Tabla 5 anterior son valores de probabilidad máximos, que comparados con los resultados de la Tabla 4 se distancia considerablemente.

Este ejercicio permite ver como la Red de Petri hace un escaneo de todo lo que pasa en el sistema, teniendo de esta forma la ventaja de ver en forma detallada y de una manera más precisa los datos de interés, como son los tiempo de cada cliente y los tamaños del sistema en cada instante, ya que está basada en una simulación en el tiempo y no en la probabilidad de lo que pueda pasar.

CAPÍTULO 8.
APLICACIÓN DE ANÁLISIS DE LÍNEA DE ESPERA A TRAVÉS DE RDP

En este capítulo se muestra la forma de hacer un análisis de una línea de espera a través de una simulación por Redes de Petri en un sistema de vida real. Esta simulación será comparada con los resultados de la Teoría de Colas y también con los datos reales por a través de un registro de frecuencias y porcentajes para concluir que modelo de análisis se acerca más a los resultados reales.

En el análisis se ponen en práctica los conocimientos registrados en este documento que son necesarios para determinar el funcionamiento de un sistema en el cual intervienen colas. Este trabajo se llevó a cabo en las cajas rápidas dentro de un importante almacén de cadena y para el análisis en este caso se siguieron los siguientes pasos: se identificaron los diferentes componentes del sistema y su funcionamiento básico, luego se llevó a cabo la recolección y organización de los datos; después de esto se procede a buscar la distribución de probabilidad de las variables de interés (tiempo entre llegadas y tiempos de servicio) para a partir de allí encontrar el modelo adecuado de teoría de colas.

Las medidas de eficiencia se obtuvieron de un archivo en Excel para facilitar su cálculo, éstas fueron al final del ejercicio con los resultados obtenidos de la simulación con las Redes de Petri.

La simulación del sistema a través de la Red de Petri fue hecha de forma muy detallada con los datos de tiempos de llegada de los clientes, se evaluó el estado del sistema en cada disparo de las transiciones y a partir de allí se obtuvieron los resultados para las medidas de eficiencia.

8.1. METODOLOGÍA DE ANÁLISIS

8.1.1. Identificación y descripción de los componentes

Los componentes de un sistema de colas como ya se sabe bien, son importantes para entender los aspectos básicos y funcionamiento del sistema, para el caso del almacén de cadena se obtuvieron los siguientes componentes:

- **Fuente de entrada:** se considera infinita; cualquier persona puede ingresar a la cola sin restricción de capacidad.
- **La cola o línea de espera:** el sistema cuenta con una única fila o cola para la atención de los clientes.
- **Mecanismo de servicio:** el sistema funciona con tres cajas de servicio, ya que durante el tiempo de análisis la cuarta caja no estuvo en funcionamiento.
- **El sistema de la cola:** la cola es de tipo FIFO, eso quiere decir que el primero en llegar es el primero en ser atendido.

8.1.2. Descripción del trabajo

Antes de empezar con cualquier análisis, es primordial tener la información adecuada, esto significa tener claro las variables a estudiar como lo son: tiempos entre una llegada y otra y los tiempos de servicio. Para lo anterior, se prepararon tres personas: una encargada de tomar el tiempo que demora un cliente en entrar en la fila, otra para tomar el tiempo que tarda cada persona antes de ser atendida y una más para tomar el tiempo de servicio. Esta información se depuró y se organizó para así obtener las variables con las cuales se trabajó.

Cabe mencionar que la recolección de datos se realizó para tres días diferentes y se llegó a la conclusión de que el sistema se comportaba de manera similar en los tres casos. Los datos se tomaron con la ayuda de un cronómetro para la disminuir los errores, con lo cual se obtuvieron los siguientes datos definidos en las tablas 6,7 y 8.

Tabla 6. Muestra de los tiempos recolectados para el día 1.

Día 1							
Cliente	**Minuto entrada**	**Tasa llegada [min/cliente]**	**Tiem servicio [min/cliente]**	**Cliente**	**Minuto entrada**	**Tasa llegada [min/cliente]**	**Tiem servicio [min/cliente]**
1	0.53	0.53	0.58	20	10.38	0.28	2.74
2	0.91	0.38	2.69	21	11.96	1.58	3.94
3	1.56	0.65	0.16	22	13.27	1.30	0.95
4	1.76	0.21	2.49	23	16.27	3.00	1.52
5	2.37	0.61	3.51	24	17.04	0.77	1.48
6	2.53	0.16	3.17	25	18.55	1.51	1.98
7	3.35	0.82	2.37	26	19.16	0.62	1.30
8	4.08	0.73	4.55	27	19.84	0.67	1.99
9	4.76	0.68	1.75	28	20.14	0.30	1.79
10	5.71	0.96	1.79	29	21.28	1.15	1.67
11	5.80	0.09	1.48	30	22.21	0.92	0.53
12	6.52	0.72	1.62	31	22.34	0.14	0.75
13	7.09	0.57	0.83	32	23.82	1.48	2.59
14	7.45	0.36	1.83	33	23.86	0.04	0.53
15	8.29	0.85	1.87	34	24.70	0.84	1.23
16	8.53	0.24	0.86	35	24.81	0.11	1.17
17	8.73	0.20	1.51	36	25.19	0.38	0.36
18	9.23	0.50	2.97	37	25.69	0.50	2.52
19	10.10	0.88	0.95	38	25.83	0.14	2.18

Tabla 7. Muestra de los tiempos recolectados para el día 2.

Día 2							
Cliente	**Minuto entrada**	**Tasa llegada [min/cliente]**	**Tiem servicio [min/cliente]**	**Cliente**	**Minuto entrada**	**Tasa llegada [min/cliente]**	**Tiem servicio [min/cliente]**
1	0.59	0.59	1.56	20	14.83	1.84	1.52
2	0.77	0.18	0.67	21	16.65	1.82	6.35
3	1.92	1.15	2.41	22	19.00	2.35	1.27
4	2.16	0.24	0.24	23	19.39	0.39	1.70
5	2.84	0.68	1.54	24	19.76	0.37	4.24
6	3.17	0.33	0.83	25	20.53	0.77	3.16
7	3.37	0.20	0.98	26	21.31	0.78	2.93
8	3.46	0.09	0.46	27	21.57	0.26	3.99
9	4.17	0.71	4.13	28	21.71	0.14	3.68
10	4.67	0.50	0.04	29	22.18	0.47	0.68
11	5.83	1.16	0.51	30	22.61	0.43	1.29
12	6.32	0.49	5.67	31	22.76	0.15	0.65
13	6.83	0.51	1.45	32	23.12	0.36	0.22
14	8.16	1.33	2.23	33	23.29	0.17	0.39
15	8.45	0.29	0.98	34	24.34	1.05	0.35
16	10.52	2.07	2.89	35	25.35	1.01	0.36
17	11.07	0.55	1.04	36	25.67	0.32	0.97
18	12.20	1.13	1.16	37	26.17	0.50	0.44
19	12.99	0.79	2.17	38	26.48	0.31	1.99

Tabla 8. Muestra de tiempos recolectados para el día 3.

Día 3							
Cliente	Minuto entrada	Tasa llegada [min/cliente]	Tiem servicio [min/cliente]	Cliente	Minuto entrada	Tasa llegada [min/cliente]	Tiem servicio [min/cliente]
1	1.96	1.96	2.19	20	17.54	1.65	4.95
2	3.70	1.73	2.90	21	17.88	0.34	0.72
3	4.96	1.27	2.63	22	19.70	1.82	1.73
4	5.75	0.79	0.22	23	20.28	0.58	0.03
5	6.99	1.24	0.67	24	20.41	0.13	1.73
6	8.10	1.11	0.62	25	21.06	0.65	0.91
7	8.54	0.43	3.21	26	21.41	0.34	3.55
8	10.08	1.55	2.95	27	21.55	0.14	1.83
9	11.42	1.34	4.86	28	22.00	0.45	3.47
10	11.75	0.32	0.17	29	22.50	0.50	0.53
11	11.85	0.11	0.66	30	23.66	1.16	0.26
12	12.04	0.18	1.11	31	24.12	0.47	1.99
13	12.43	0.39	2.20	32	24.88	0.75	0.72
14	13.15	0.72	3.41	33	25.24	0.37	2.33
15	13.56	0.41	0.90	34	25.62	0.38	1.20
16	14.03	0.46	0.02	35	25.97	0.34	0.35
17	14.80	0.77	3.36	36	26.11	0.14	0.98
18	15.39	0.59	2.58	37	26.55	0.44	2.08
19	15.89	0.50	1.76	38	26.67	0.12	2.29

De ahora en adelante los datos de la Tabla 6, los llamaremos datos reales ya que al final se hará una comparación general de estos datos con los obtenidos de la simulación en RdP y por Teoría de Colas.

De los datos reales obtenidos se obtuvieron los siguientes resultados.

Tabla 9. Tasa de llegadas, tasa de servicio y factor de utilización para los datos reales.

ρ	λ	μ
0.8696	1.47	0.56

Tabla 10. Frecuencias de los datos para las medidas de eficiencia en el caso real.

L		Lq		W		Wq	
n	Frecuencia de n	n	Frecuencia de n	t [min]	Frecuencia de t	t [min]	frecuencia de t
0	4%	0	50%	0.00	0%	0.00	39%
1	13%	1	11%	1.06	0%	0.47	13%
2	18%	2	9%	2.12	18%	0.94	5%
3	14%	3	9%	3.18	34%	1.41	8%
4	11%	4	9%	4.23	42%	1.88	8%
5	9%	5	8%	5.29	5%	2.35	11%
6	18%	6	4%	6.35	5%	2.81	16%
7	8%						
8	4%						

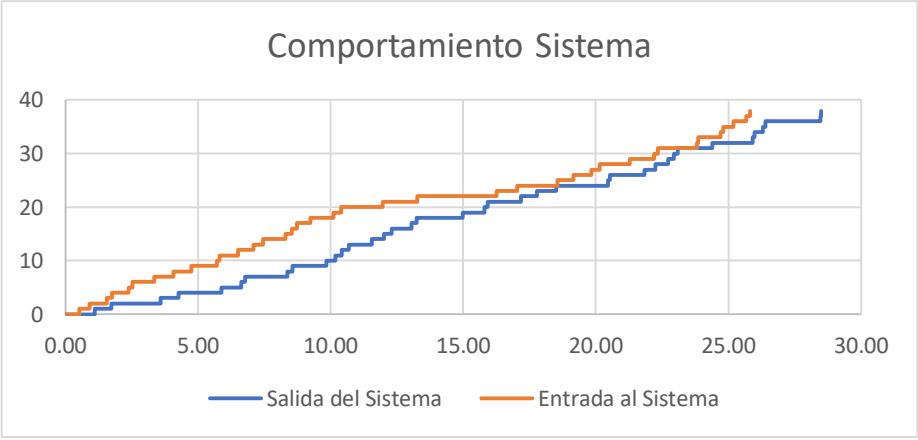


Figura 18. Comportamiento del sistema real.

La gráfica presentada en la Figura 18 muestra cómo el sistema se comporta en el tiempo, dando información de los clientes que entran (línea naranja) y salen del sistema (línea azul). Se observa que alrededor del minuto 9 se encuentra el número máximo de clientes en el sistema que es de 9 clientes; en los minutos 18.55 y 23.82 las dos líneas se encuentran haciendo que la diferencia sea cero, lo cual significa que en esos instantes no hay clientes en el sistema; la horizontal entre la línea

naranja y la línea azul representa el tiempo de permanencia en el sistema, en donde el máximo valor que se puede extraer de la gráfica es de 4.6 min.

8.1.3. Análisis de distribución de datos

En el presenta numeral se hace el análisis del comportamiento de los datos recolectados únicamente para la Tabla 6. El análisis se realiza con la herramienta de análisis de datos STATGRAPHICS®. Primero se analizan los tiempos entre llegadas y luego los tiempos de servicio.

Tiempos entre llegadas

A continuación se evidencia el análisis hecho por STATGRAPHICS® para el ajuste probabilístico de los datos. Se ajustaron los datos a una distribución exponencial como se muestra en la Figura 19, estos datos cuentan con una media de 0.68. Los resultados del análisis fueron los siguientes:

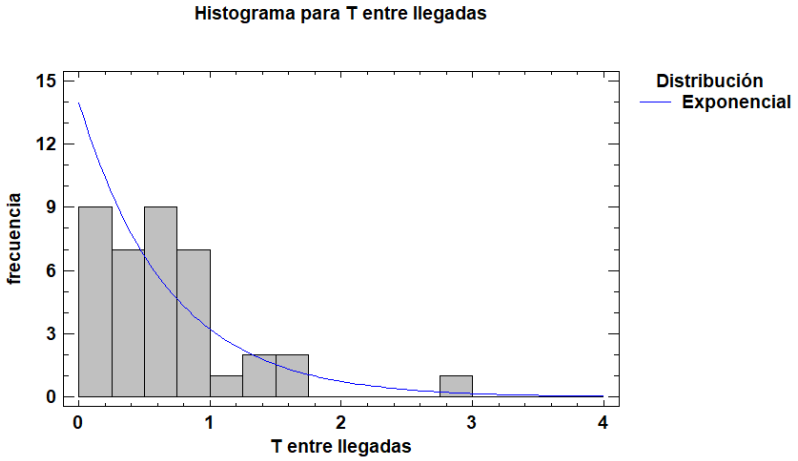


Figura 19. Histograma y ajuste para el tiempo entre llegadas.

Para justificar mejor el ajuste a una distribución exponencial, el programa además realiza una prueba de Kolmogorov-Smirnov para determinar si el tiempo entre llegadas se ajusta realmente a una distribución exponencial. De allí se obtiene el valor-p, el cual es el mínimo nivel de significancia en el cual es rechazada la hipótesis de que un conjunto de datos siga una distribución especificada, por lo tanto, el valor-p es la alternativa más simple a la conclusión de rechazar o no rechazar una hipótesis de ajuste.

$$\text{Valor} - \text{p} = 0.347963$$

Como el valor-P fue mayor que 0.05, no se puede rechazar la hipótesis y se concluye que hay evidencias suficientes para pensar que los datos provienen de

que los datos de tiempo entre llegadas provienen de una distribución exponencial con 95% de confianza.

Tiempos de servicio

Al mirar el histograma de los tiempos de servicio se puede notar que también sigue una distribución exponencial de media 1.798, como se ve a en la Figura 20:

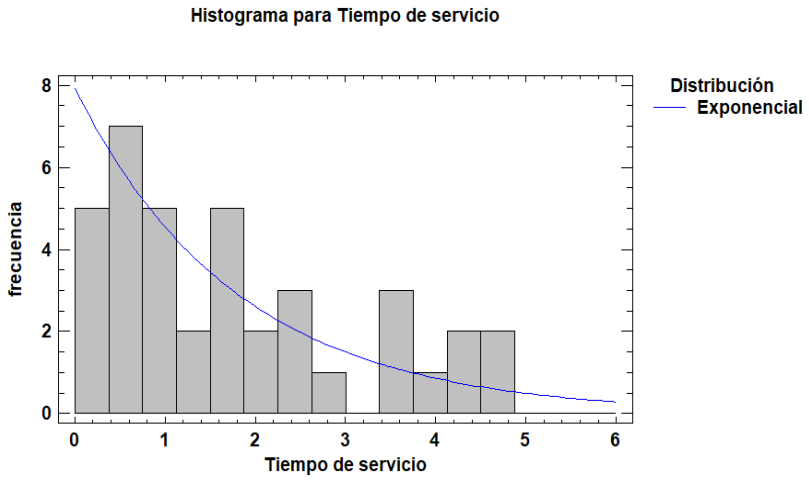


Figura 20. Histograma y ajuste para el tiempo de servicio.

Para verificar los resultados obtenidos, es necesario realizar una prueba de bondad de ajuste de igual forma que el caso de los tiempos entre llegadas y así corroborar que la distribución exponencial si se puede ajustar a los datos analizados.

$$\text{Valor} - \text{p} = 0.734039$$

Debido a que el valor-P más pequeño de las pruebas realizadas es mayor a 0.05, no se puede rechazar la idea de que el tiempo de servicio proviene de una distribución exponencial con un 95% de confianza.

Al tener claro las distribuciones que siguen las variables de interés y entendiendo a través de los componentes el número de servidores, se establece el tipo de modelo adecuado para el análisis por medio de Teoría de Colas. Gracias a lo dicho en la descripción del trabajo y específicamente en el mecanismo de servicio, se puede decir que el sistema cuenta con tres servidores. Tanto los tiempos entre llegadas y tiempos de servicio siguen una distribución exponencial, por lo tanto, su notación es de la siguiente forma:

$$\text{M/M/3}$$

8.1.4. Análisis por Teoría de Colas

Por medio de un archivo en Excel diseñado para aplicar las fórmulas (20), (21), (22) y (23) ya vistas de un sistema M/M/C, se obtienen los siguientes resultados para el cálculo de las medidas de eficiencia, como se muestra en Tabla 11:

Tabla 11. Medidas de eficiencia para el análisis por Teoría de Colas.

Lq [clientes]	Wq [minutos]	W [minutos]	L [clientes]	Fac. utilización (ρ)
5	3.5	5.3	8	0.8696

La Tabla 11 muestra que la probabilidad es que se encuentren 5 clientes en la cola y que la mayor cantidad de clientes es de 8, por otra parte, la probabilidad de tiempo en la cola es de 3.5 minutos y el mayor tiempo de espera de un cliente es de 5.3 minutos

8.1.5. Solución por Redes de Petri

La RdP que describe este sistema se puede apreciar en la Figura 21, está diseñada con una transición de entrada al sistema (T1) con distribución exponencial de parámetro λ igual a 1.47, tres transiciones de entrada a los servidores con disparos instantáneos (T2, T4 y T6), tres transiciones de salida del sistema (T3, T5 y T7) con μ de 0.56, un lugar (P2) con marcas “infinitas” para el límite de la cola, un lugar (P1) donde se almacenan las marcas de la cola, tres lugares (P3, P5 y P7) donde pasan las marcas a ser servidas y por último tres lugares (P4, P6 y P8) que limitan el servicio a una marca.

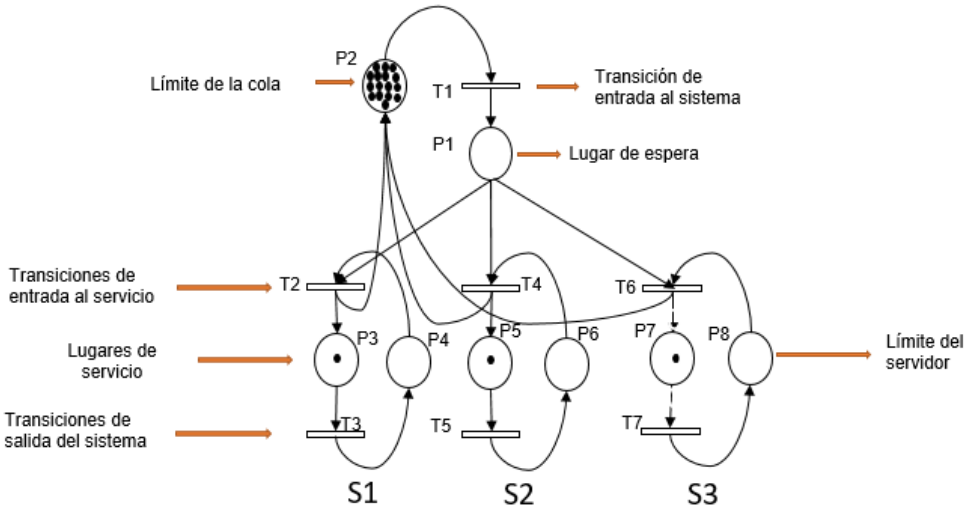
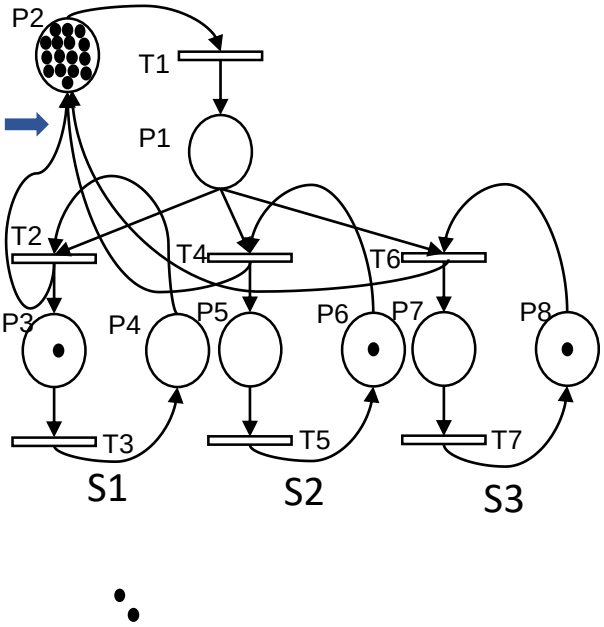


Figura 21. Representación en RdP para el sistema de la aplicación.

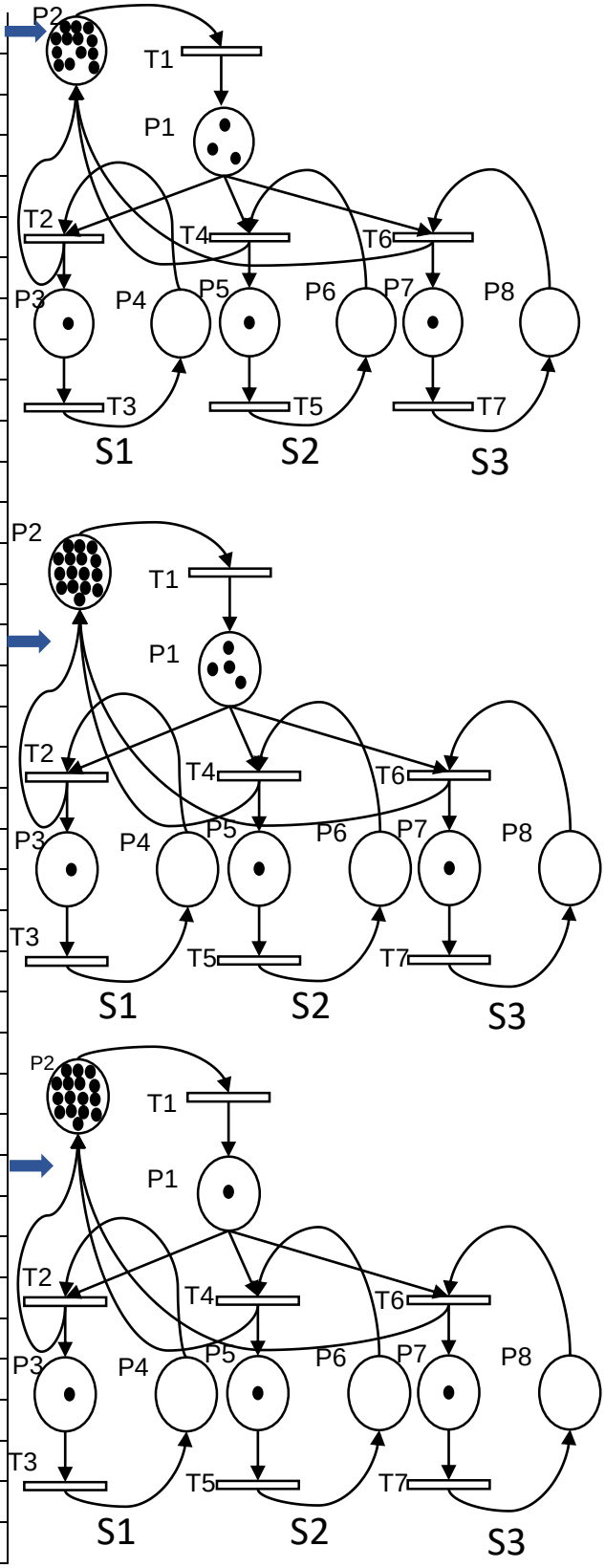
➢ Simulación

La simulación consiste en describir la cantidad de disparos a lo largo del tiempo de análisis. La Figura 22 está compuesta por una tabla y cinco subgráficos, la tabla se compone por ocho columnas, la primera columna muestra el tiempo en el que las transiciones se disparan, las otras siete columnas representan los disparos de las transiciones a medida que transcurre el tiempo, las transiciones T1, T3, T5 y T7 tienen un comportamiento aleatorio siguiendo una distribución exponencial, mientras que las transiciones T2, T4 y T6 poseen un comportamiento instantáneo. A la derecha y en conjunto con la tabla, se puede visualizar cinco subgráficos que describen el estado del sistema en distintos instantes de tiempo, es decir, la cantidad de clientes en la cola, la cantidad de clientes siendo atendidos en los servidores y la disponibilidad de cada uno de los servidores.

Tiempo [min]	DISPARO TRANSICIONES						
	T1 Exp	T2 Ins	T3 Exp	T4 Ins	T5 Exp	T6 Ins	T7 Exp
0.53	1	1	-	-	-	-	-
0.91	2	-	-	1	-	-	-
1.00	-	-	1	-	-	-	-
1.56	3	-	-	-	-	1	-
1.76	4	2	-	-	-	-	-
1.79	-	-	-	-	1	-	-
2.37	5	-	-	2	-	-	-
2.53	6	-	-	-	-	-	-
2.56	-	3	2	-	-	-	-
3.35	7	-	-	-	-	-	-
3.65	-	4	3	-	-	-	-
4.08	8	-	-	-	-	-	-
4.16	-	-	-	-	-	2	1
4.22	-	-	-	-	2	-	-
4.33	-	-	-	-	-	-	2
4.76	9	-	-	3	-	-	-
5.71	10	-	-	-	-	3	-
5.80	11	-	-	-	-	-	-
6.03	-	5	4	-	-	-	-
6.52	12	-	-	-	-	-	-
7.09	13	-	-	-	-	-	-



7.45	14	-	-	-	-	-	-
7.89	-	-	-	-	-	4	3
8.29	15	-	-	-	-	-	-
8.53	16	-	-	-	-	-	-
8.73	17	-	-	-	-	-	-
8.82	-	-	-	4	3	-	-
9.15	-	-	-	-	-	5	4
9.23	18	-	-	-	-	-	-
9.32	-	-	-	-	-	6	5
10.10	19	-	-	-	-	-	-
10.38	20	-	-	-	-	-	-
10.54	-	6	5	-	-	-	-
11.82	-	-	-	-	-	7	6
11.96	21	-	-	-	-	-	-
13.27	22	-	-	-	-	-	-
13.47	-	-	-	5	4	-	-
13.64	-	-	-	-	-	8	7
14.19	-	7	6	-	-	-	-
14.36	-	8	7	-	-	-	-
14.68	-	-	-	6	5	-	-
16.06	-	-	8	-	-	-	-
16.27	23	9	-	-	-	-	-
16.55	-	-	-	-	6	-	-
17.04	24	-	-	7	-	-	-
17.16	-	-	-	-	-	-	8
18.55	25	-	-	-	-	9	-
18.90	-	-	-	-	7	-	-
19.16	26	-	-	8	-	-	-
19.84	27	-	-	-	-	-	-
20.14	28	-	-	-	-	-	-
20.21	-	-	-	9	8	-	-
20.55	-	10	9	-	-	-	-
20.63	-	-	-	-	9	-	-
21.26	-	-	10	-	-	-	-
21.28	29	-	-	10	-	-	-
22.21	30	11	-	-	-	-	-
22.34	31	-	-	-	-	-	-
23.28	-	-	-	-	-	10	9



23.67	-	-	-	-	-	-	10
23.82	32	-	-	-	-	11	-
23.86	33	-	-	-	-	-	-
23.97	-	-	-	-	-	12	11
24.70	34	-	-	-	-	-	-
24.81	35	-	-	-	-	-	-
24.94	-	-	-	-	-	13	12
24.99	-	-	-	11	10	-	-
25.08	-	-	11	-	-	-	-
25.19	36	12	-	-	-	-	-
25.29	-	-	-	-	11	-	-
25.35	-	-	-	-	-	-	13
25.69	37	-	-	12	-	-	-
25.83	38	-	-	-	-	14	-
25.86	-	-	12	-	-	-	-
26.28	-	-	-	-	-	-	14
27.57	-	-	-	-	12	-	-

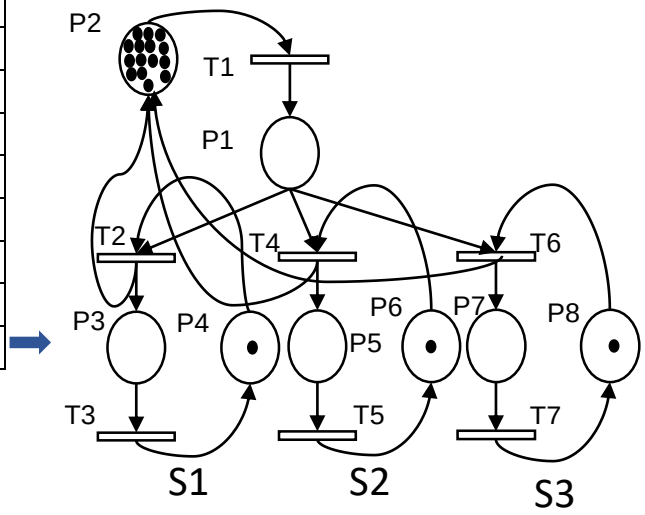


Figura 22. Simulación en RdP ejemplo aplicativo

De la tabla de la Figura 22 se puede extraer que la transición T1 se disparó 38 veces, lo que significa que entraron 38 clientes al sistema; las transiciones T2, T4 se dispararon doce veces cada una, lo que quiere decir que doce clientes ingresaron al servidor uno (S1) y doce clientes ingresaron al servidor dos (S2), posteriormente estos clientes fueron atendidos, por esto las transiciones T3 y T5 también se dispararon doce veces cada una; las transiciones T6 y T7 se dispararon 14 veces cada una, lo que significa que al servidor tres (S3) ingresaron, fueron atendidos y salieron del sistema 14 clientes.

La Tabla 12 deja ver los resultados para las medidas de eficiencia W y Wq para cada uno de los clientes que ingresa al sistema y posteriormente sale. Entrega información de cuánto tiempo permanece un cliente en la cola y en el sistema.

Tabla 12. Resultados de tiempos de la simulación en RdP.

TIEMPOS						
CLIENTE	**Wq**	**W**		**CLIENTE**	**Wq**	**W**
1	0.00	0.47		20	3.80	3.97
2	0.00	0.89		21	2.39	4.09
3	0.00	2.61		22	1.42	3.29
4	0.00	0.79		23	0.00	4.28
5	0.00	1.85		24	0.00	1.86
6	0.03	1.12		25	0.00	4.73
7	0.30	2.68		26	0.00	1.05
8	0.08	0.25		27	0.38	0.80
9	0.00	4.06		28	0.41	1.12
10	0.00	2.18		29	0.00	3.71
11	0.23	4.73		30	0.00	2.87
12	1.37	2.63		31	0.94	1.32
13	1.73	6.39		32	0.00	0.15
14	1.70	1.87		33	0.11	1.08
15	1.03	3.53		34	0.24	0.65
16	2.01	5.66		35	0.19	0.48
17	3.09	4.91		36	0.00	0.67
18	4.25	5.46		37	0.00	1.89
19	3.54	7.06		38	0.00	0.45

A continuación se muestra la Tabla 13, la cual entrega información acerca del tamaño de la cola y del sistema de cada para cada instante de tiempo analizado.

Tabla 13. Resultados de tamaño de cola y de sistema para la RdP.

TAMAÑO DEL SISTEMA									
Tiempo [min]	**Lq**	**L**		**Tiempo [min]**	**Lq**	**L**		**Tiempo [min]**	**Lq**	**L**
0.53	-	2		8.82	4	7		20.55	-	3
0.91	-	2		9.15	3	6		20.63	-	2
1.00	-	1		9.23	4	7		21.26	-	1
1.56	-	2		9.32	3	6		21.28	-	2
1.76	-	3		10.10	4	7		22.21	-	3
1.79	-	2		10.38	5	8		22.34	1	4
2.37	-	3		10.54	4	7		23.28	-	3
2.53	1	4		11.82	3	6		23.67	-	2
2.56	-	3		11.96	4	7		23.82	-	3
3.35	1	4		13.27	5	8		23.86	1	4
3.65	-	3		13.47	4	7		23.97	-	3
4.08	1	4		13.64	3	6		24.70	1	4
4.16	-	3		14.19	2	5		24.81	2	5

4.22	-	2		14.36	1	4		24.94	1	4
4.33	-	1		14.68	-	3		24.99	-	3
4.76	-	2		16.06	-	2		25.08	-	2
5.71	-	3		16.27	-	3		25.19	-	3
5.80	1	4		16.55	-	2		25.29	-	2
6.03	-	3		17.04	-	3		25.35	-	1
6.52	1	4		17.16	-	2		25.69	-	2
7.09	2	5		18.55	-	3		25.83	-	3
7.45	3	6		18.90	-	2		25.86	-	2
7.89	2	5		19.16	-	3		26.28	-	1
8.29	3	6		19.84	1	4		27.57	-	0
8.53	4	7		20.14	2	5				
8.73	5	8		20.21	1	4				

- **Resumen del sistema**

La Tabla 14 muestra un resumen general de los valores máximos de las medidas de eficiencia.

Tabla 14. Resumen del sistema.

Max W	Max Wq	Max L	Max Lq	ρ	λ	μ
7.06	4.25	8	5	87%	1.47	0.56

En la Tabla 15 se muestran los porcentajes que de los valores que más se repiten para el tamaño del sistema, el tiempo en el sistema, tamaño de la cola y tiempo en la cola.

Tabla 15. Probabilidades de las medidas de eficiencia para la simulación en RdP.

Tamaño del sistema		**Tiempo en el sistema**		**Tamaño de la cola**		**Tiempo en la cola**	
n clientes	**p(x=n)**	**s [min]**	**P(t=s)**	**n clientes**	**p(x=n)**	**s [min]**	**P(t=s)**
0	1%	0.00	0%	**0**	**57%**	**0.00**	**37%**
1	7%	**1.18**	**37%**	1	16%	0.71	32%
2	22%	2.35	16%	2	7%	1.42	11%
3	**26%**	3.53	13%	3	8%	2.12	8%
4	16%	4.70	16%	4	9%	2.83	3%
5	7%	5.88	13%	5	4%	3.54	5%
6	8%	7.06	5%			4.25	5%
7	9%						
8	4%						

De la Tabla 15 se puede decir que el sistema en la mayoría de las veces hay aproximadamente tres personas, el tiempo de espera en el sistema que más

frecuencia tiene es de aproximadamente 1.18 minutos, el tiempo de espera en cola que más se repite es de cero minutos y el número de personas en la cola que más frecuencia tuvo fue cero clientes.

8.1.6. Comparaciones Real vs Teoría de Colas y RdP

Tabla 16. Tabla de comparaciones para las medidas de eficiencia entre los datos reales, teoría de colas y RdP.

MEDIDA DE EFICIENCIA	REAL	TEORÍA DE COLAS	RdP
L	El tamaño del sistema estuvo en 2 clientes la mayor parte del tiempo	La mayor probabilidad para este sistema es que se encuentre con 8 clientes	La simulación muestra que el sistema tiene la mayor probabilidad de tener entre 2 y 3 clientes en el sistema
Lq	La cola estuvo en 0 la mayor parte del tiempo	La mayor probabilidad es que la cola sea de 5 clientes.	La simulación muestra que el sistema tiene la mayor probabilidad de 0 clientes en la cola
W	El tiempo en el sistema estuvo en 2.12 minutos para la mayor cantidad de clientes	La mayor probabilidad es que los clientes tarden en el sistema 5.3 minutos	La mayor probabilidad es que los clientes estén 1.18 minutos en el sistema
Wq	El tiempo en cola estuvo para la mayor cantidad de clientes en 0 minutos	La mayor probabilidad es que los clientes tarden 3.5 minuto en la cola	La mayor probabilidad es que los clientes tarden 0 minutos en la cola

8.1.7. Análisis de costos

Para el administrador del sistema de la línea de espera, físicamente solo hay un valor que puede modular, el cual es la tasa de servicio del sistema, este parámetro se puede controlar sumando o quitando servidores. La tasa de entrada se puede controlar accediendo directamente a los clientes, por ejemplo, asignar citas para que lleguen a determinada hora; para este caso solo se tiene en cuenta la modificación del sistema como tal bajo la presión del parámetro λ (Lambda) ya que aún si se modifica, el sistema de igual forma debe analizarse y modificarse según la presión causada por la llegada de clientes.

Esta presión en el sistema se conoce como factor de utilización del sistema (ρ) como se vio anteriormente aplicado en los ejemplos, recordando que:

$$\rho = \frac{\lambda}{c * \mu}$$

La ecuación anterior indica, que si se aumenta la cantidad de servidores del sistema disminuye el factor de utilización y consecuente a esto disminuye el tiempo de espera y el tamaño de la cola y si se disminuye los servidores, sucede lo contrario. Cambiar el factor de utilización acarrea costos, por ello, es un valor a criterio del analista según el comportamiento que espera del sistema. Para este caso se evalúa lo que sucede con el factor de utilización para el ejemplo aplicativo si se cambia la cantidad de servidores como se ve en la Tabla 17.

Tabla 17. Análisis del factor de utilización para cierta cantidad de servidores.

Servidores	Lq [clientes]	Wq [min]	W [min]	L [clientes]	Fac. utilización (ρ)
2	--	--	--	--	1.3043
3	**5**	**3.5**	**5.3**	**8**	**0.8696**
4	1	0.6	2.4	3	0.6522
5	0	0.2	2.0	3	0.5217

De la tabla anterior, se puede decir que no es viable colocar dos servidores ya que los resultados muestran que el sistema se satura y no sería posible disminuir la cola, ya que se observa que el factor de utilización está por encima del 100%, lo que significa que el sistema se encuentra saturado; para 4 servidores, el factor de utilización baja hasta un 65% y para 5 servidores baja hasta 52%, estos valores son considerables y se tienen en cuenta a la hora de reajustar el sistema ya que muestran el flujo de los clientes y mejora la satisfacción de los mismos, pero a su vez, implican uso subutilizado de los servidores. Para el caso de 3 servidores, se puede decir que la opción ya que presenta un factor de utilización del 87%, lo que quiere decir que los servidores se ocupan la mayor parte del tiempo, además el sistema no se encuentra en saturación.

Por esto es importante tener una metodología más asertiva que la teoría de colas como lo permiten las RdP y tener la certeza de la cantidad de servidores que se necesitan para que el sistema funcione de una manera deseada, encontrando un equilibrio entre la satisfacción de los clientes y el costo de brindar esa satisfacción.

CAPÍTULO 9.

CONCLUSIONES

- La Teoría de Colas es una herramienta muy útil porque permite realizar predicciones acerca de tiempos y tamaños del sistema para así tomar decisiones que lo optimicen de acuerdo al número de servidores, pero aun así, presenta una restricción para el análisis de sistemas saturados ($\rho > 1$).
- Las Redes de Petri gracias a su sistema lógico de flujo permiten diseñar de forma adecuada un sistema de líneas de espera por complejo que sea, además no posee limitaciones en cuanto a saturación o líneas de espera independiente dentro de un mismo sistema; dado lo anterior, las RdP permiten analizar sistemas complejos con tasas de llegadas y servicio con diferentes distribuciones.
- La principal limitante para los sistemas que poseen un solo servidor es la capacidad para prestar el servicio sin que la fila crezca de forma descontrolada, por esto, existen sistemas con múltiples servidores que ayudan a atender mayor cantidad de clientes en menos tiempo y controlando el tamaño de la fila.
- Los modelos de colas determinísticos no son adecuados para la simulación de sistemas de atención al público, ya que allí los clientes no presentan un comportamiento conocido de manera exacta, los clientes normalmente llegan de manera aleatoria a una fila y los servidores también atienden siguiendo ese mismo comportamiento, estas situaciones solo se puede modelar a través de sistemas probabilísticos, los cuales presentan un modelamiento acertado en sistemas de colas donde clientes sean personas.
- Para un sistema de atención al público, el número de servidores condiciona el análisis de los parámetros que miden la eficiencia como los tiempos de espera y el tamaño del sistema, pero además, su costo-beneficio también se ve directamente afectado por el número de servidores, si la cantidad de servidores no es suficiente para suplir la demanda de clientes, provoca que los mismos servidores trabajen en sobrecarga y además que los clientes esperen demasiado tiempo para ser atendidos, reflejándose como pérdidas por demora en el servicio, por el contrario, si hay demasiados servidores los clientes serán atendidos más rápido pero puede haber servidores desocupados demasiado tiempo, por lo tanto, puede haber pérdidas por subutilización de servidores.

BIBLIOGRAFÍA

[1] QMATIC, «QMATIC,» [En línea]. Available: https://www.qmatic.com. [Último acceso: 19 Julio 2018].

[2] M. E. C. Ernesto, «ANÁLISIS DE REDES DE COLAS MODELADAS CON TIEMPOS ENTRE LLEGADAS EXPONENCIALES E HÍPER ERLANG PARA LA ASIGNACIÓN EFICIENTE DE LOS RECURSOS,» Pontificia Universidad Javeriana, Bogotá, 2009.

[3] S. J. M. &. S. A. Fonollosa Joan B, MÉTODOS CUANTITATIVOS DE ORGANIZACIÓN INDUSTRIAL II, Catalunya: Edicions UPC, 2005.

[4] P. Larrea, CALIDAD DEL SERVICIO: DEL MARKETING A LA ESTRATEGIA, Madrid: Ediciones Díaz de Santos, 1991.

[5] J. C. S. T. M. &. N. H. R. Joselito Medina Marín, «APLICACIÓN DE REDES DE PETRI EN LA MODELACIÓN DE SISTEMAS DE EVENTOS DISCRETOS,» Universidad Autónoma del Estado de Hidalgo, 2013.

[6] W. Guaito, «MODELOS DE SIMULACIÓN DE EVENTOS DISCRETOS Y DE PROCESOS CONTINUOS,» [En línea]. Available: http://dinamica-de-sistemas.com/revista/0608o.htm. [Último acceso: 21 Febrero 2019].

[7] C. P. F. Rodriguez, «ANÁLISIS DE UN SISTEMA DE EVENTOS DISCRETOS MEDIANTE REDES DE PETRI,» Escuela Politécnica Nacional, Quito, 2006.

[8] I. E. Ángeles, «METODOLOGÍA PARA LA MODELACIÓN, SIMULACIÓN Y ANÁLISIS DE PROCESOS DE MANUFACTURA UTILIZANDO REDES DE PETRI,» Tecnológico de Monterrey, Monterrey, 2005.

[9] M. A. P. Cayuela, «Webs: Universidad de Murcia (colast6),» [En línea]. Available: https://webs.um.es/mpulido/miwiki/lib/exe/fetch.php?id=amio&cache=cache&media=wiki:colast6.pdf.. [Último acceso: 7 Febrero 2019].

[10] P. S. Y.-M. &. J. A. H. Gutierrez, «UNA INTRODUCCIÓN AMABLE A LA TEORÍA DE COLAS,» 2018.

[11] RECORDANDO A ERLANG: UN BREVE PASEO POR LA TEORÍA DE COLAS, Barcelona: Publicació electrónica de divulgació del Departament de Matemàtiques de Universitat Autònoma de Barcelona, 2009.

[12] J. M. H. Mendoza, «AMBIENTE VISUAL DE SIMULACIÓN Y ANÁLISIS DE REDES DE COLAS,» Instituto Politécnico Nacional, México D.F., 2007.

[13] J. P. G. Sabater, «APLICANDO TEORÍA DE COLAS EN DIRECCIÓN DE OPERACIONES,» Universidad Politécnica de Valencia, Valencia, 2015.

[14] G. A. L. M. PATIÑO PASQUEL JHINNER ALEJANDRO, «BENEFICIOS DEL ANÁLISIS DEL TIEMPO DE ESPERA EN FILA GENERAL O PREFERENCIAL PARA EL ACUEDUCTO Y ALCANTARILLADO DE POPAYÁN S.A E.S.P,» Universidad del Cauca, Popayán, 2018.

[15] M. A. P. Cayuela, «Webs: Universidad de Murcia (colast5),» [En línea]. Available: https://webs.um.es/mpulido/miwiki/lib/exe/fetch.php?id=amio&cache=cache&media=wiki:colast5.pdf.. [Último acceso: 10 Febrero 2019].

[16] [En línea]. Available: https://www.um.es/or/ampliacion/node5.html. [Último acceso: 10 Mayo 2019].

[17] S. J. M. A. &. J. M. V. Montoya, *APLICACIÓN DE LAS REDES DE PETRI AL CONTROL DE UN ASCENSOR AUTOMÁTICO,* Pereira: Universidad Tecnológica de Pereira, 2007.

[18] M. Granada, *REDES DE PETRI: DEFINICIÓN, FORMALIZACIÓN Y EJECUCIÓN,* Universidad de Cantabria, 2012.

[19] Á. Á. O. G. &. C. G. L. MAURICIO HOLGUÍN LONDOÑO, AUTOMATISMOS INDUSTRIALES, Pereira: Universidad Tecnológica de Pereira, 2008.

[20] A. Sabiguero, «NOMENCLATURA Y DEFINICIONES BÁSICAS DE REDES DE PETRI,» Universidad de la República-Uruguay.

[21] «Portal Estadística Aplicada,» [En línea]. Available: http://www.estadistica.net/INVESTIGACION/CADENAS-MARKOV.pdf. [Último acceso: 21 Abril 2019].

[22] C. L. L. &. J. M. B. Carlos Luis Flores, «TEORÍA DE COLAS Y SU APLICACIÓN AL SISTEMA BANCARIO,» Universidad de El Salvador, 2009.